



元照出版提供 請勿公開散布

刑事程序中人工智慧於 風險評估上的應用

李 榮 耕*

要 目

- | | |
|------------------------|---------------------|
| 壹、前 言 | (二)「再犯」概念的界定 |
| 貳、人工智慧及再犯風險評估 | (三)資料中影響再犯的因素為何 |
| 一、人工智慧的簡述 | (四)資料的取捨及修改 |
| 二、科技於刑事程序中的應用 | (五)建構預測模型 |
| 三、人工智慧於司法中的應用 | (六)風險因子是否及如何加權 |
| 四、風險評估系統於刑事司法中的應用 | (七)個案數據的取得及輸入 |
| 五、刑事程序中的風險評(預)估(及其問題) | (八)建構風險評估工具 |
| 參、風險評估系統的建置 | (九)風險群組的建立 |
| 一、風險評估系統的概述 | (十)正確性(accuracy)的審查 |
| 二、系統建置上的問題：「客觀」的風險評估系統 | 肆、運作上不透明／黑盒的疑慮 |
| (一)資料的選擇及蒐集 | 一、現象及其成因 |
| | 二、相關疑慮及知悉原因的需要 |
| | (一)錯誤或是不合理之處的即時發現 |
| | (二)透明性及訴訟上的防禦權 |

DOI : 10.53106/102398202022030168003

* 臺北大學法律學院教授，美國印地安納大學布魯明頓校區摩爾法學院法學博士。

投稿日期：一一〇年一月二十二日；接受刊登日期：一一〇年七月二十五日

責任校對：鍾淑婷



伍、風險評估系統在使用上應依循的原則

一、透明性的要求

(一) 相關資訊及事項的公開

(二) 應對外揭露採用風險評估系統的標準及程序

(三) 系統建置者易傾向於不公開相關資料

(四) 透明化便足以正當化風險評估系統的使用（？）

二、當責性

(一) 概念

(二) 外部及民主參與

(三) 使用時的應告知事項

(四) 獨立的查核機制及事後的評估程序

三、不應作為唯一或決定性因素

四、刑罰的目的及再犯風險的評估

五、自動化依賴的預防

陸、結論



摘 要

由於人工智慧技術日趨成熟，部分國家（如美國）已經開始在刑事程序中使用風險評估系統，預測特定人可能存有的風險，決定是否拘禁其人身自由或釋放條件。不過，建置風險評估系統的每一個環節中都有著人的主觀及價值判斷。再者，因為人工智慧技術的本質，風險評估系統會有黑盒運作的疑慮。是故，我們建議，若是要在刑事程序中使用風險評估系統，應該要遵循透明性的要求。此外，系統的使用也必須要具備其當責性，以確保系統達到其所設定的目標、結果的正確性及保障人民的人身自由等權益。再者，系統的評估結果不應作為刑事程序中決定的決定性因素，且應審視系統的建置目的與程序的規範目的是否一致。最後，風險評估系統的使用必須要避免自動化依賴的情形發生。

關鍵詞：人工智慧、風險、風險評估系統、透明、當責、自動化依賴



壹、前言

近年來，人工智慧（artificial intelligence, AI）成了耳熟能詳的名詞，許多公共政策或議題的討論，也都圍繞在人工智慧打轉。事實上，人工智慧的概念早在一九五五年就已經為學者所提出¹，只是到了最近才因為硬體的運算能力及計算的關鍵技術的問世，才又再度躍上了舞臺，成為各界關注的焦點²。相關技術的成熟，使得人工智慧不再受限於人們已知的知識範疇，更邁向了廣泛應用，如語音辨識、語言翻譯、圖形辨識、自動駕駛、醫療診斷及經濟或金融預測分析等等³。也正因為技術已達門檻，人工智慧借助於巨量資料（big data）及機器學習（machine learning），有能力針對特定範圍內的事物進行預測，如預判消費者可能感興趣的商品、可能對某商品感興趣的族群、金融商品的漲跌，甚至是新研發的藥物的效果等等。國家機關在形成決定時，也都已經可以看到人工智慧的實際應用，其中包括了以人工智慧建置風險評估系統（Risk Assessment System），刑事程序中的法官或是其他官員依據或參考系統的運算結果，作成裁判或是處遇（如是否釋放或是羈押、釋放條件、刑度、假釋或假釋條件等）。乍看之下，這樣的作法使得相關的決定更為客觀、公正，且降低或避免了人所可能會有的主觀偏見，但事實上是否如此，實有深入研究的必要。

這一篇論文擬討論及分析的議題是，在刑事程序中導入人工智

¹ John McCarthy et al., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (Aug. 31, 1955), 27(4) AI MAGAZINE 12, 12-14 (2006).

² 關於人工智慧的發展，可以參照李榮耕，初探刑事程序法的人工智慧應用，載：人工智慧相關法律議題芻議，頁122-123，2018年11月。

³ 針對人工智慧在不同時期的技術取向，可以參照劉靜怡，人工智慧潛在倫理與法律議題鳥瞰與初步分析，載：人工智慧相關法律議題芻議，頁3-4，2018年11月。



慧所建置的風險評估系統，預測特定人可能具有的風險，風險評估系統的建置過程是否客觀中立，抑或是存有人主觀決定在內。再者，外界是否能夠瞭解系統的應用，即時發現其中可能的錯誤，即時修正，相對人在程序中能否為有意義的防禦。最後在面對這些挑戰時，有哪些可能的解決方案及應該要依循的原則。為要深入探究此一議題，以下擬先說明人工智慧已經如何在司法實務展露頭角，以人工智慧所建置的風險評估系統可能如何應用於刑事司法程序中，以及其所引發的不同意見。接著，這篇論文會分析風險評估系統的建置過程、其可能受有的人的影響，以及系統在運作上所會有的疑慮，並在最後針對前述的問題，提出將來要在刑事司法程序中導入風險評估系統，所應依循的原則及限制。由於無論在司法實務、法院裁判及學理上，美國在此一議題都已經累積有大量的經驗及文獻，故以下會以美國法制及學界研究的成果作為比較法制研究上的參考及借鏡。

貳、人工智慧及再犯風險評估

在過往，人工智慧只是存在於理論或實驗室中的技術，但由於電腦計算能力的大幅進步及相關科學的快速發展，在人們生活中到處可見人工智慧的應用。亦即，人工智慧技術已經逐漸成熟，達到了能夠商業應用的程度，部分國家甚至已經在刑事程序中使用此一技術。

一、人工智慧的簡述

人工智慧已經是可以朗朗上口的詞彙，但其實質意涵，卻仍難以定義及掌握。其中的原因在於，「人工」很容易就可以理解，約莫意指非自然，人力所為或所造，但相對地，什麼是「智慧」或

元照出版提供 請勿公開散布

「有智慧的行為」就很難明確定界定其概念⁴。不過，關於人工智慧，一般多能夠接受的描述是，利用程式，使電腦具有部分接近於人或是動物認知能力的技術⁵。

具體地來說，人工智慧，包含了機器學習（machine learning）、語音辨識（speech recognition）、自然語言處理（natural language processing）及圖像辨識（image recognition）等技術。其中，機器學習的功能在於，從巨量的資料中，發現或形成規律，並針對其他的事件，作成預測或評估⁶。由於電腦有著強大的計算能力，所以機器學習可以在短時間內，處理極為大量的資料，甚至在某些事物上作成比人更為正確的決定，例如在棋類競賽或是電玩競技，電腦便已經能夠贏過棋王或是世界冠軍⁷。由於類似的電腦技術已經有著相當的成熟度，所以其已經走出了實驗室、競賽或遊戲的場域，應用在人們日常的生活當中，其中包括了在刑事程序中的許多階段裡。乍看之下，這樣的描述或許使人詫異，但在過往，刑事程序就已經在許多階段或事項中運用各樣的科技，協助司法、執法或其他

⁴ 關於「人工智慧」的難以定義，可以參照李榮耕，同註2，頁122；邱文聰，初探人工智慧中的個資保護發展趨勢與潛在的反歧視難題，載：人工智慧相關法律議題芻議，頁153，2018年11月。

⁵ Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399, 404 (2017).

⁶ Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1, 6 (2019).

⁷ See AlphaZero AI Beats Champion Chess Program After Teaching Itself in Four Hours, *The Guardian* (Dec. 7, 2017), available at <https://www.theguardian.com/technology/2017/dec/07/alphazero-google-deepmind-ai-beats-champion-program-teaching-itself-to-play-four-hours> (last visited: 2019.11.29); Dan Garisto, Google AI Beats Top Human Players at Strategy Game StarCraft II, *Nature* (Oct. 30, 2019), available at <https://www.nature.com/articles/d41586-019-03298-6> (last visited: 2019.11.29).

元照出版提供 請勿公開散布。
官員作成決定。

二、科技於刑事程序中的應用

在過去，在刑事程序中，多是由人（司法警察官員、檢察官、法官及其他刑事司法官員）來評價證據及各樣資訊，認定事實，作成決定。不過，時至今日，由於科技飛快的進步，越來越多的科學儀器或是技術進入了司法體系中，成為了官員們在判斷時的重要協助或依據。舉例來說，以往要確認一個人的供述是否真實，是透過觀察陳述時的神情、交互詰問、具結，以及陳述的內容與其他證據是否一致來決定，但是在對於測謊原理、測（說）謊生理及心理、測謊設備及技術有的更為深入的研究及更多的掌握後，在刑事司法中，開始有了「測謊」（polygraph）技術的使用，以各樣的儀器量定或偵測受測者的生理反應，透過人緊張時的身體跡徵變化（心跳、血壓、瞳孔及出汗等），確認其陳述是否為真⁸。雖然在美國司法實務，多數的法院都認為，測謊的結果沒有證據能力，無論是檢察官或被告都不能以測謊證明實體事實⁹，因為測謊並不可信，且會不當地左右了陪審團對於證詞的信用性的判斷¹⁰，但是，在判決後的程序，刑事司法機關仍大量地使用測謊，以為各樣決定

⁸ See KEN ALDER, THE LIE DETECTORS: THE HISTORY OF AN AMERICAN OBSESSION 108 (2007).

⁹ See *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

¹⁰ See, e.g., *United States v. Scheffer*, 523 U.S. 303, 307 (1998); *United States v. Alexander*, 526 F.2d 161, 170 (8th Cir. 1975). 關於測謊，國內的文獻可以參照曾春僑，儀器測謊與常見疑義，在野法潮，46期，頁74-79，2020年7月；黃虹霞，測謊報告與證據能力、證明力——一個執業律師的看法，萬國法律，149期，頁31-36，2006年10月；王兆鵬，重新思維測謊之證據能力，月旦法學雜誌，135期，頁137-145，2006年8月；吳巡龍，科學證據與測謊的證據能力，月旦法學教室，38期，頁95-105，2005年12月。

元照出版提供 請勿公開散布

的參考。例如，美國的緩刑及假釋部門便經常要求受判決人，特別是在妨害性自主案件，接受測謊，作為相關處置（如治療）的條件，在撤銷緩刑或假釋的聽證中，也經常以測謊報告為證據，證明受判決人的陳述是否真實¹¹。

除了測謊之外，司法警察機關在執法時，也常會使用到許多儀器，如呼氣酒精檢測儀（breathalyzer, or breath blood tester）。詳細地來說，過去警察多只能憑藉其主觀的觀察（神情、眼睛是否泛有血絲或是否渾身酒氣等）、本人的陳述及證詞，判斷一個人是否醉酒。雖然警察可以使用酒醉測試或是平衡動作測驗（走直線、單腳維持一定時間的平衡及畫同心圓等），但不少人即使喝了酒，還是可以輕易地通過，因此在個案中誤判者，不在少數¹²。於一九五〇年代，呼氣酒精檢測儀問世後，其測試結果成為了案件中確認特定人是否酒醉的關鍵，該儀器可以說完全取代了過往的執法方式，除了在少數的個案外，實務的運作上，已經不會再以要求走直線及主觀判斷的方式來判斷一個人是否酒醉¹³。

測謊或是呼氣酒精檢測儀只是在刑事司法裡運用新興技術或是儀器的一個很小的例子，但卻足以窺見科技在國家機關執法中，已經扮演了越來越吃重的角色。因著電腦技術的飛快進步及人工智慧已經具可應用性，司法，乃至於刑事司法中，已經開始運用人工智

¹¹ See, e.g., Douglas C. Maloney, *Lies, Damn Lies, and Polygraphs: The Problematic Role of Polygraphs in Postconviction Sex Offender Treatment (PCSOT)*, 84 TEMP. L. REV. 903, 907-08 (2012). Also *J.B. v. New Jersey Parole Bd.* 131 A.3d 413 (2016).

¹² See Andrea Roth, *The Uneasy Case for Marijuana as Chemical Impairment Under a Science-Based Jurisprudence of Dangerousness*, 103 CAL. L. REV. 841, 850 (2015).

¹³ Joseph Dorlaque, *New Machines Get Drunks off the Road*, POPULAR SCI. MONTHLY 165, 166 (1955).

元照出版提供 請勿公開散布

慧，幫助，甚至是代替官員作成各樣原本單純是由人所作成的決定。

三、人工智慧於司法中的應用

人工智慧的應用，除了在日常生活或商業活動中使用外，也開始可見於司法程序。二〇一五年，史丹佛大學的學生Joshua Browder建置了DoNotPay聊天機器人，可以在詢問幾個簡單的問題後，產出一份訴狀，讓使用者用以申訴違規停車的事件。之後，DoNotPay的服務更擴及到了航空公司的航班遲延及各樣的小額訴訟¹⁴。IBM則是利用其知名的人工智慧「華生」（Watson）打造了AI律師Ross¹⁵。Ross甫推出，紐約的Baker & Hostetler事務所馬上便「聘用」了它，負責破產事件。事務所的律師或是職員以一般的用語查詢後，Ross在數秒鐘之內就可以找到相關的法規及過往判決的立場。在過去，這是律師及法務助理可能必須要耗費數十，甚至是數百個小時才能完成的工作¹⁶。

不只是律師，人工智慧也開始在訴訟程序中扮演仲裁者的角色。愛沙尼亞於二〇一七年，命其資訊長著手設計AI法官（AI judge, or robot judge），用以處理爭訟金額小於七千歐元（約合新臺幣二十四萬五千元）的事件。雙方當事人各自將訴狀及相關資料上傳到系統，AI法官就會作成裁決。當事人不服者，可以向一般的法官提起上訴。愛沙尼亞的目標是，藉由AI法官的導入可以適度地

¹⁴ DoNotPay, available at <https://donotpay.com> (last visited: 2021.01.21).

¹⁵ Ross, available at <https://www.rossintelligence.com/> (last visited: 2020.05.04).

¹⁶ Anthony Sills, Ross and Watson Tackle the Law, available at <https://www.ibm.com/blogs/watson/2016/01/ross-and-watson-tackle-the-law/> (last visited: 2020.05.04).

減輕司法體系的既有負擔，降低法官及法院人員所要處理的案件量¹⁷。雖然整個制度仍在建置及研議當中，但愛沙尼亞很有可能是第一個賦予電腦程式作成司法裁判權力的國家。在美國也已經有部分的州，在刑事案件的審判中，使用演算法，提供給法院建議的量刑¹⁸。

¹⁷ Eric Niiler, Can AI Be a Fair Judge in Court? Estonia Thinks So, *Wired* (Mar. 25, 2019), available at <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/> (last visited: 2019.09.16).

¹⁸ E.g., *State v. Loomis*, 371 Wis.2d 235 (2016). 本案發生於威斯康辛州，事實約為，在一起駕車槍擊（drive-by shooting）案件中，被告是駕車者，並起訴了逃逸未遂、未經所有人同意便駕駛他人車輛及持有槍械等罪名。被告否認涉入該起案件，但針對其中兩個案件認罪，檢察官放棄追訴其他三個案件，但是在協商中載明（read in）該三案件可以成為量刑的依據。雙方當事人對於事實的部分都沒有爭執（*Loomis*, 371 Wis.2d at 244.）。被告認罪後，巡迴法院（circuit court）命觀護人（probation officer）作成「量刑前調查報告」（presentence investigation report, PSI），其中包括了COMPAS的風險評估結果（*Loomis*, 371 Wis.2d at 244-45.）。COMPAS是一套由Northpointe公司所設計的風險評估系統，以被告的犯罪資料及面談為基礎，判定被告與那個族群有著類似的背景或條件者，若是被釋放，再犯他罪的可能性，用以幫助矯正機關決定安置或治療等處遇（*Loomis*, 371 Wis.2d at 245-46.）。COMPAS的報告指出，被告有著極高風險值，同時說明了應該如何使用該評估報告、可能的誤用、應以報告確認就相對人有益的處置及應留心的風險因子等事項。COMPAS的報告特別強調，法院不應該以風險評估的結果判定刑度的輕重及是否得受緩刑宣告（*Loomis*, 371 Wis.2d at 246.）。在量刑程序中，檢察官主張法院在決定刑度時，應將COMPAS報告納入考慮，而法院也的確參酌了COMPAS報告及其他因素（*Loomis*, 371 Wis.2d at 246-47.）。被告並未接受法院的作法，主張法院在量刑時考慮了風險評估報告的作法，違反了正當法律程序的要求，因為：(1)量刑並未基於正確的資訊，被告無從檢視COMPAS的正確性；(2)被告未受有個別化的量刑；以及(3)法院在量刑時，不當地將性別納入考量（*Loomis*, 371 Wis.2d at 248, 251-52.）。被告進而向威斯康辛州最高法院提起上訴。威斯康辛州最高法院未接受被告的主張。威斯康辛州最高法院判定，在限制使用範圍的前提下，法院並不是不能在量刑程序中使用COMPAS之類的風險評估系統。為遵循正當法律程序，矯正機關及過往的判決都已經說明了在量刑程序中，使用風險評估程式所應依循的原則及其限

元照出版提供 請勿公開散布

四、風險評估系統於刑事司法中的應用

如同前段所述，在部分國家中，已經在司法程序中使用人工智慧，幫助官員在個案中作成決定。以人工智慧進行風險評估，也是相當典型的例子。

詳細地來說，風險評估指的是，利用過往的資料，評估具有類似背景的特定人，能夠應傳喚到場、再犯或其他情事的可能性高低¹⁹。在刑事程序的許多階段中，都會評估犯罪嫌疑人、被告或受判決人的再犯風險。在羈押審查程序中，釋放與否及釋放條件為何，能否宣告緩刑，以及是否准予假釋的程序中，都會審酌相對人再犯的可能性高低²⁰。

關於再犯的風險評估，政策上多希望能夠有較為客觀的評估或預測，避免流於官員的主觀的臆測或是恣意的獨斷。舉例來說，有部分國家會以犯罪紀錄作為量刑的最低刑度的標準²¹，或是三振法案（three strike law）的制度²²。我國刑法也有類似於三振法案的規

制，如風險評估的結果不能作為釋放被告與否的決定性因素等等，在本案中的下級審法院也都已經遵循，所以威斯康辛州最高法院駁回被告的上訴，維持原裁判（*Loomis*, 371 Wis.2d at 253-56.）。

¹⁹ John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725, 1729 (2018).

²⁰ 如法務部，收容人處遇透明化、假釋透明化，<https://www.moj.gov.tw/dl-25932-555a3b36c1334097a95822f950f1595a.html>，最後瀏覽日：2020年5月6日（指出在判斷假釋與否時，再犯風險是考量的因素之一）；法務部，法務部新聞稿（2005年1月7日），<https://www.moj.gov.tw/dl-26449-d7a5e410a9df406d8d75b076cfa9acc4.html>，最後瀏覽日：2020年5月4日（再犯的預防是考量緩刑的因素之一）。

²¹ BERNARD E. HARCOURT, *AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE* 91 (2007).

²² E.g., Jessica M. Eaglin, *The Drug Court Paradigm*, 53 AM. CRIM. L. REV. 595, 601, 615-16 (2016).

定。亦即，「受徒刑之執行而有悛悔實據者，無期徒刑逾二十五年，有期徒刑逾二分之一、累犯逾三分之二，由監獄報請法務部，得許假釋出獄。」（第77條第1項）但是「犯最輕本刑五年以上有期徒刑之罪之累犯，於假釋期間，受徒刑之執行完畢，或一部之執行而赦免後，五年以內故意再犯最輕本刑為五年以上有期徒刑之罪者。」（同條第2項第2款）也就是以一定期間內的再犯罪，作為高再犯風險的認定，排除假釋規定的適用²³。這都是希望藉由明確的條文，在一定程度上限縮司法官員主觀判斷的空間，使得再犯風險的評估或人身自由的限制與否更具可預測性及一致性。

由於電腦技術的可觀進步，開始有了使用演算法或是人工智慧進行風險評估的倡議。詳細地來說，在過去，就特定人的再犯可能性的判斷，多是以書面表格、資料或列點式的文件，計算被告的再犯可能性高低，但現在則有主張認為應改用既有的資料，透過深度學習，訓練人工智慧，評估特定人的再犯風險²⁴。美國律師公會（American Bar Association, ABA）便曾建議，於刑事司法體系中，應採用風險評估工具（risk assessment tools）²⁵。其中的理由

²³ 本條的修正理由為：「……(二)對於屢犯重罪之受刑人，因其對刑罰痛苦之感受度低，尤其犯最輕刑本五年以上重罪累犯之受刑人，其已依第一項規定（執行逾三分之二）獲假釋之待遇，猶不知悔悟，於1.假釋期間、2.徒刑執行完畢或赦免後五年內再犯最輕刑本五年以上之罪，顯見刑罰教化功能對其已無效益，為社會之安全，酌採前開美國『三振法案』之精神，限制此類受刑人假釋之機會應有其必要性，爰於第二項第二款增訂之。」（底線為作者所加）

²⁴ See Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 638 (2017).

²⁵ American Bar Association, Criminal Justice Section, State Policy Implementation Project 18, available at http://www.americanbar.org/content/dam/aba/administrative/criminal_justice/spip_handouts.authcheckdam.pdf (last visited: 2019.09.19).



在於，實證研究顯示，自由刑不利於低風險的受判決人，因為他們與中度或高度風險的受判決人在監獄一起服刑的結果，並無法降低其再犯率，反而是提高²⁶。亦即，若是無法準確地判斷哪些受判決人的風險較低，使其接受監獄監禁以外的處遇，並無助於他們的社會復歸（*rehabilitation*），反而是增加了對於一般社會大眾的危害。是故，ABA主張應導入風險評估系統的使用，提高犯罪預測的正確性，進而降低再犯率及促進公共安全。

在不同的國家的刑事司法程序中，都已經各種類型的風險評估系統或技術的應用²⁷。這些風險評估系統因著開發的目的、所使用的資料及技術的差異，有著不同的功能，有些系統是針對一般性的再犯預測，有些則是針對特定罪名（如妨害性自主）的再犯預測。舉例來說，加拿大心理學家及學者所設計的Violent Risk Appraisal Guide (VRAG)便是以12個變數，預測特定人再犯的可能性²⁸，而私人企業Multi-Health Systems所開發的LSI-R，除了預測再犯風險外，也可以評估特定條件是否有助於降低特定人再犯罪的可能性²⁹。風險評估工具一開始主要是應用在緩刑及假釋程序中³⁰，但之後法

²⁶ *Id.* at 19.

²⁷ 這些風險預測或是評估技術，有使用於警察勤務預警勤務，劃定犯罪熱區者，也有使用於審前及量刑程序，建議是否拘禁被告及釋放條件者。See, e.g., Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327, 341-48 (2015). 這篇論文說明了人工智慧及演算法如何應用於合理懷疑的判斷。See also Sandra G. Mayson, *Dangerous Defendants*, 127 YALE L.J. 490, 507-17 (2018). 這篇論文的内容介紹了風險評估程式如何應用於量刑。

²⁸ Melissa Hamilton, *Back to the Future: The Influence of Criminal History on Risk Assessments*, 20 BERKELEY J. CRIM. L. 75, 93 (2015).

²⁹ Michael Tonry, *Legal and Ethical Issues in the Prediction of Recidivism*, 26 FED. SENT'G REP. 167, 172 (2014).

³⁰ Pamela M. Casey et al., National Center for State Courts (NCSC), Using Offender

院也開始在量刑程序中使用這類的工具，用以協助法官，判斷被告再犯的可能性高低，決定是否監禁被告或是釋放條件³¹。從另一個角度來說，以人工智慧所建置的風險評估系統也可以避免法官或司法官員其僅憑個人主觀，甚至是偏見或不當目的（如減輕監所容留人數的壓力等），決定被告的監禁、釋放或是刑度³²。由於依據客觀的證據，而不是恣意決定是否監禁特定人及釋放條件，是刑事司法體系中的重要利益，所以，美國許多州都已經引入了風險評估系統，提供法官更多的資訊，以期作成更為正確及妥適的量刑決定，如威斯康辛州（Wisconsin）、肯德基州（Kentucky）³³、俄亥俄州（Ohio）³⁴、賓州（Pennsylvania）³⁵、亞利桑納州（Arizona）³⁶及

Risk and Needs Assessment Information at Sentencing: Guidance for Courts from a National Working Group 1 (2011), available at <https://www.ncsc.org/~media/Microsites/Files/CSI/RNA%20Guide%20Final.ashx> (last visited: 2019.09.19). See also John Monahan & Jennifer L. Skeem, *Risk Assessment in Criminal Sentencing*, 12 ANN. REV. CLINICAL PSYCHOL. 489, 493-94 (2016).

³¹ See, e.g., National Center for State Courts (NCSC), Conference of Chief Justices and Conference of State Court Administrators, Resolution 7: In Support of the Guiding Principles on Using Risk and Needs Assessment Information in the Sentencing Process (2011), available at <https://www.ncsc.org/~media/Microsites/Files/CSI/Resolution-7.ashx> (last visited: 2019.09.19); John Monahan, *A Jurisprudence of Risk Assessment: Forecasting Harm Among Prisoners, Predators, and Patients*, 92 VA. L. REV. 391, 405-06 (2006). Also Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 805 (2014).

³² See Algorithms in the Criminal Justice System, Elec. Privacy Info. Ctr., available at <https://epic.org/algorithmic-transparency/crim-justice/> (last visited: 2019.12.03). 這份調查報告中指出，截至2019年9月為止，美國聯邦及49個州，都在刑事程序中的不同階段，使用了風險評估工具。

³³ Ky.Rev.Stat. Ann. § 532.007(3)(a) (2016).

³⁴ Ohio Rev.Code Ann. § 5120.114(A)(1)-(3) (2015-16).

奧克拉荷馬州 (Oklahoma)³⁷ 等等。

前述的風險評估系統，絕大多數都是由私人企業所推出³⁸。其中包括了知名的Northpointe所開發的COMPAS³⁹，以及Multi-Health Systems的LSI-R⁴⁰。只有少部分的程式是由政府所研發，如維吉尼亞州 (Virginia) 及賓州⁴¹。還有部分是由公私實體組織所開發，如加拿大心理學家及學者所設計的Violent Risk Appraisal Guide (VRAG)⁴²，以及辛辛那提大學的矯治研究中心 (University of Cincinnati Corrections Institute) 開發了俄亥俄州風險評估系統 (Ohio Risk Assessment System, ORAS)⁴³。

³⁵ 42 PA. Cons.Stat. § 2154.7(a) (2016).

³⁶ Ariz. Code of Judicial Admin. § 6-201.01(J)(3) (2016).

³⁷ Okla. Stat. tit. 22, § 988.18(B) (2016).

³⁸ See Monahan & Skeem, *supra* note 30, at 499.

³⁹ Northpointe, Inc., Practitioner's Guide to COMPAS Core (2015), *available at* http://www.northpointeinc.com/downloads/compas/Practitioners-Guide-COMPAS-Core-_031915.pdf (last visited: 2019.12.20).

⁴⁰ Pamela M. Casey et al., Nat'l Ctr. for State Courts, Offender Risk & Needs Assessment Instruments: A Primer for Courts app. A-38 (2014), *available at* https://www.ncsc.org/~media/Microsites/Files/CSI/BJA%20RNA%20Final%20Report_Combined%20Files%208-22-14.ashx (last visited: 2019.12.20).

⁴¹ Richard P. Kern & Meredith Farrar-Owens, *Sentencing Guidelines with Integrated Offender Risk Assessment*, 16 FED. SENT'G REP. 165, 166 (2004); 42 Pa. Stat. and Cons. Stat. Ann. § 2154.7 (West Supp. 2017).

⁴² See Vernon L. Quinsey et al., *Violence Risk Appraisal Guide (VRAG)*, in *VIOLENT OFFENDERS: APPRAISING AND MANAGING* 283, 283-86 (Vernon L. Quinsey, Grant T. Harris, Marnie E. Rice & Catherine A. Cormier eds., 2d ed. 2006).

⁴³ Ohio Risk Assessment System, National Institute of Correction, *available at* <https://nicic.gov/ohio-risk-assessment-system> (last visited: 2019.12.25).

五、刑事程序中的風險評（預）估（及其問題）

對於特定人的再犯風險的評估，影響到刑事程序中的諸多環節。在這一些階段中，判斷的一致性，相當地重要，過去實務運作常為人所詬病的方面，就是在類似案件中，評估結果的不一致⁴⁴。是故，有認為風險評估系統的使用，可以幫助官員形成更為客觀、精準且一致的決定，促進風險評估結果的一致性、妥適性及正確⁴⁵。相對地，反對者則認為，風險評估系統的使用有著公平性，甚至是違憲的疑慮⁴⁶。再者，如果用以訓練人工智慧的資料中有偏見，系統的運算結果也就無可避免地存在偏見，甚至是會放大偏見的效果⁴⁷。人工智慧在風險評估上的表現，也常是被質疑的地方。有研究顯示，人工智慧在這一個領域的應用上，並沒有較高的正確率，甚至不比一個沒有刑事司法知識的人來得好⁴⁸。是故，沒有必要大費周

⁴⁴ 如王宏舜，無感的司改？法官信賴度始終偏低，聯合報，2018年2月13日，A4版（文中敘及，司法院院長許宗力表示，量刑的差異是民怨的來源之一）。另外，可以參照王宏舜，看盡法庭眾生 民眾最在意量刑，聯合報，2019年6月18日，A8版。

⁴⁵ See, e.g., Jordan Hyatt et al., *Reform in Motion: The Promise and Perils of Incorporating Risk Assessments and Cost-Benefit Analysis into Pennsylvania Sentencing*, 49 DUQ. L. REV. 707, 713 (2011).

⁴⁶ Dawinder S. Sidhu, *Moneyball Sentencing*, 56 B.C. L. REV. 671, 703-05 (2015). 這篇文章立論的理由在於，風險評估系統可能將性別或種族作為考量的因素，所以有其妥適性上的疑慮。

⁴⁷ See, e.g., ANDREW GUTHRIE FERGUSON, *THE RISE OF BIG DATA POLICING* 131-32 (2017).

⁴⁸ See, e.g., Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, 4(1) SCI. ADVANCES 1 (Jan. 2018), available at <https://advances.sciencemag.org/content/4/1/eaao5580/tab-pdf> (last visited: 2019.12.03). 這份研究顯示，用以預測再犯率的COMPAS的表現，與一個沒有刑事司法經驗或知識的人相差無幾。



章地在刑事司法程序中引入人工智慧，建置風險評估系統。最後，人工智慧或是演算法的編碼、結構、變數及權重等事項，外界幾乎都無從得知，更遑論審查其是否正確或是考量了不當的因素⁴⁹。整體來說，對於在刑事程序中使用風險評估系統，似乎可以解決司法中長久以來，為人詬病的部分問題，但電腦程式或系統的導入，仍然受有各樣的質疑，擔憂其未能真正因應司法體系所面臨的困境，甚至是引發了更多新的爭議。

這樣的憂慮，並非無的放矢，杞人憂天。學者們的研究顯示，風險評估系統的建置及使用中，的確可能存在著為人所忽視或輕忽等問題。為要更瞭解在刑事程序中使用風險評估系統所可能帶來的助益及衝擊，以下擬先說明如何建置風險評估系統，在建置過程中可能影響評估結果的各個主要事項，以及在其中可能受有人的價值判斷及主觀影響的部分。

參、風險評估系統的建置

刑事程序中的諸多環節都涉及到了各種不同風險的評估。在過往，這些事項都是靠法官等司法官員、心理或精神鑑定專業人士或社工師等人員來判斷，但近來，由於電腦運算技術的長足進步，許多國家已經開始在這些程序中導入人工智慧，由其計算特定人的風險值，並據以作成各樣關係到人身自由的決定。從外觀上看來，電腦技術的運用，似乎較為公正客觀，不會受到人的主觀因素的影響，但事實上，並非如此。風險評估系統的建置過程中，在每一個重要的環節，都有人的主觀想法或價值判斷在其中。

⁴⁹ Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1349-50 (2018).

元照出版提供 請勿公開散布

一、風險評估系統的概述

乍看之下，風險評估系統是單純的科學，其計算、運作及產出的結果，都是從客觀的資料、演算法及電腦程式所運算而來，不會有人的想法、主觀或甚至是偏見在內，所以可以獲得較為公平且一致的風險的評估。然而，事實上，運用人工智慧對於特定人再犯風險評估的結果，還是有著人的價值判斷在內，甚至絲毫不小於傳統的判斷方式。

簡單地來說，風險評估系統是以過往的資料為基礎，分析歸納其中的特定事項（變數），以人工智慧所建構的預測模式，判斷特定人將來再犯罪的可能性。不過，電腦程式或是演算法都是為了某個特定的目的而存在，所以其會依據人們所輸入的資料，執行操作者所下達的指令，忠實地反應人們對於事實的假定及價值的選擇⁵⁰。就算是看似單純的資料，也都有著人的主觀意向在內，其中包括選擇什麼樣的資料、納入資料中的什麼數值、決定數值與運算結果間的關聯，以及統計模型等等，無不都涉及到人的判斷⁵¹。換言之，風險評估系統看似是透過人工智慧或演算法，針對特定人的再犯風險產出客觀的結果，但其實系統本身還是受到一定的限制，難以突破「人」所可能帶來的缺點或是極限，甚至可能會在系統中考量不當的因素，如性別、種族及階級等，也因而使得少數或弱勢族群立於極為，甚至是更為不利的處境⁵²。

⁵⁰ DAVID A. MINDELL, OUR ROBOTS, OURSELVES: ROBOTICS AND THE MYTHS OF AUTONOMY 10 (2015).

⁵¹ Editorial Itiel Dror, *The Ambition to Be Scientific: Human Expert Performance and Objectivity*, 53 SCI. & JUSTICE 81, 81 (2013). See also Robin Marantz Henig, *Death by Robot*, N.Y. TIMES, Jan. 9, 2015, at MM16.

⁵² See Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 10-16 (2014).



風險評估系統的建置，可以約略分為兩個階段。首先是決定的是要使用哪些資料、從哪些地方取得這些資料、定義再犯的概念，以及要以資料中的哪一些事項來判斷再犯可能性的高低。之後，再依據前述的結論，寫成程式碼，演算再犯的可能性。第二個階段涉及到如何呈現運算的結果，以供司法官員於具體個案中應用，判斷再犯風險的高低。兩個階段中的核心關鍵問題，都關乎到價值判斷，諸如什麼是再犯？或是什麼樣的風險是社會大眾所能（必須要）忍受的？由於風險是社會對於個別事件的集體感受，不是單純可以量化的數據，也因此，在刑事程序中無論是否使用風險評估系統，都避免不了都會以某種形式，做出主觀的選擇。以下擬逐一審視建置風險評估系統的每一個步驟，分析其中可能涉及價值判斷者。

二、系統建置上的問題：「客觀」的風險評估系統

風險評估系統的建置，歷經幾個關鍵的步驟。在每一個環節中，都有著人的觀念或是判斷在其中，並不是如一般人所想像，只要是由系統或是程式所運算出來的數據，就是中立或沒有偏見的結果，也因此系統對於特定人再犯的預測，並非全然客觀。

(一)資料的選擇及蒐集

風險評估系統的核心是人工智慧（或演算法），也因此，建置上的第一步是要先蒐集資料。如何蒐集訓練人工智慧所使用的資料、自何處取得、如何整合不同的資料，都會影響到系統對於再犯風險的評估結果。

詳細地來說，為建構風險評估系統，開發者多會選擇蒐集及使用起訴或是有罪判決的紀錄。其中較為普遍的，就是從政府機關所公布的各樣資料，如法院所公告的判決書，或是各樣的統計年報。



部分風險評估系統還會使用到緩刑資料。ORAS就使用到了俄亥俄州數個郡中681名被起訴的被告，包括其二〇〇六年九月至二〇〇七年二月間的緩刑資料⁵³。部分風險評估系統使用了美國東北部區域的資料（如Northpointe所開發的COMPAS）⁵⁴。針對特定州所設計的風險評估工具所使用的，必然是該州的相關資料，不過，其可能只使用了州的部分區域的資料。例如，賓州量刑委員會為研發風險評估工具所進行的研究，只參酌了全州67個郡中的4個郡⁵⁵。俄亥俄州則使用了全州88個郡中的14個郡的相關資料⁵⁶。建置風險評估系統所使用的資料選擇，會有以下幾個疑慮。

首先，僅僅使用部分區域的資料，可能會有代表性不足的問題。亦即，部分區域的資料（逮捕紀錄、起訴書、判決書或是假釋決定等）只能顯示出該區域的情形，很難反映出整體地區的情形，因為某個區域的犯罪率及類型，可能有別於其他地方（如因為是老舊社區，所以有比較多的暴力犯罪）。是故，以部分區域的資料所建置的風險評估系統，應用到整個地域時，可能就會有過度高估或低估犯罪活動，因而失準。

⁵³ Edward Latessa et al., Univ. of Cincinnati, Creation and Validation of the Ohio Risk Assessment System: Final Report 14 (2009), *available at* https://www.assessments.com/assessments_documentation/ORAS/ORAS_Final_Report_and_Validation.pdf (last visited: 2021.01.21).

⁵⁴ Northpointe, Inc., *supra* note 39, at 15.

⁵⁵ Pa. Comm'n on Sentencing, Risk/Needs Assessment Project: Interim Report 1: Review of Factors Used in Risk Assessment Instruments 8 (2011), *available at* <http://pcs.la.psu.edu/publications-and-research/research-and-evaluation-reports/risk-assessment/phase-i-reports/interim-report-1-review-of-factors-used-in-risk-assessment-instruments> (last visited: 2020.02.07).

⁵⁶ Latessa et al., *supra* note 53, at 13.



再者，資料是由人所製造，從來都不是客觀的事實的紀錄⁵⁷。特定事務的資料，只反映了其中的特定面向，而不是所有的細節。也因此，某件事物的相關資料中，只會有製作者因著特定的目的，所選擇紀錄的項目，資料中也就只會有該項目的內容。所以，為了其他的目的而使用某項資料時，就可能超過了該資料起初所製作的目的，並不完全能夠供該目的所使用。舉例來說，一個大學畢業生的成績單所紀錄的是其在大學期間的學習成果，但成績單並不完全能夠反映該大學生在畢業後的工作表現及能力。單以大學成績單來判斷一個人是否能勝任某項工作，可能就會有很大的誤差。同樣地，無論是起訴書、緩刑資料、判決書或統計年報，都有其當初製作的目的及所涵蓋的事項，幾乎不可能是為了建置風險評估系統所蒐集，也因此，這些資料不當然完全符合其需要，系統運算出來的結果，可能也就與實際情形有所出入。

此外，資料的選擇本身，也會影響到風險評估系統的運算結果，進而可能造成對於特定族群的偏見。以我國近年來的警政署所公布的統計資料為例，我國每一年約發生29萬起犯罪案件⁵⁸。其中，最高比例的犯罪嫌疑人是基層技術工及勞工或無業者，約為21%及23%⁵⁹。以教育程度來說，約25%的犯罪嫌疑人學歷在國中以下⁶⁰。但是，警察機關所認定及移送的犯罪嫌疑人，不一定會被諭知有罪判決，因為檢察官可能認為犯罪嫌疑不足而不起訴（刑訴法第252條第10款）或是起訴後，法官依檢察官提出的證據，無法

⁵⁷ Kate Crawford, *The Hidden Biases in Big Data*, HARV. BUS. REV. (Apr. 1, 2013), available at <https://hbr.org/2013/04/the-hidden-biases-in-big-data> (last visited: 2019.12.25).

⁵⁸ 內政部警政署，107年警政統計年報，2019年7月，頁34（表4）。

⁵⁹ 同前註，頁240（表8）。

⁶⁰ 內政部警政署，同註58，頁246（表8）。



獲致毋庸置疑的心證（同法第301條）。是故，如果以警政署的統計資料作為再犯罪的標準，在評估基層技術工、勞工或無業者，或是學歷在國中以下之人的再犯風險時，可能就會有歧視或是偏見的問題，因為在評估系統所使用的資料中，這一些族群在統計上有著更高的犯罪率，系統也就因而更容易判定這一些人有更可能再犯罪。但這樣的結果，不當然就是正確的（因為其最後可能未被起訴或未受有罪判決），理論上也有著違反平等原則的疑慮。在美國部分的州的研究便顯示，COMPAS對於再犯率的預測，於不同種族人間有著相當大的差異，非裔被告較容易被誤判為有再犯的高風險，白人則容易被誤判較不可能再犯⁶¹。

（二）「再犯」概念的界定

決定所使用的資料後，系統建置者接下來要定義：「什麼是『再犯』」？⁶²詳細地來說，再犯指的是某個人於將來再犯罪，而這至少涉及到幾個問題。第一，什麼是「再犯」或「再犯罪」？第二，「將來」指的是未來的多久的一段時間內？⁶³這些問題的答案，都涉及到人的主觀判斷、價值選擇及不同情境下的決定。首先，再犯罪可以是「再被判決有罪確定」，也可以是被諭知有罪判決、被起訴、被逮捕，或是緩刑或假釋被撤銷而入監服刑。再者，再犯可以是再次犯任何一種犯罪，或是再犯特定的犯罪（如重罪、暴力犯罪或是性侵害犯罪等）。至於要將再犯罪定義指的是那一種情形，就必須要由開發者，依評估系統的功能及目的來決定。

⁶¹ Loomis, 371 Wis.2d at 263. 不過，Northpointe爭執這份報告的正確性。

⁶² Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 678 (2016).

⁶³ Robert Weisberg, *Meanings and Measures of Recidivism*, 87 S. CAL. L. REV. 785, 787-88 (2014).



以現有的風險評估系統來說，多半是逮捕來定義再犯罪⁶⁴。使用逮捕認定再犯的優點在於，簡單、成本低，且容易取得。知名的ORAS及COMAPS都是以逮捕界定（再）犯罪⁶⁵。不過，也有以被起訴認定再犯罪者，如VRAG是以暴力犯罪被起訴作為再犯者。其中，暴力犯罪的概念除了參考加拿大的法規，涵蓋有或是無肢體接觸的攻擊行為外，也包含了非暴力的幼童凌虐⁶⁶。

第二，系統建置者也必須要決定，再犯指的是在犯罪後多長的一段時間內，再發生犯罪。現有的風險評估系統都有著不完全相同的設定。VRAG以十年以上的資料，預測五年內的再犯可能性⁶⁷。ORAS所預測的則是一年內的再犯率⁶⁸。COMPAS的再犯預測期間，則是兩年⁶⁹。至於為什麼是一年、兩年或是五年，原因不一，很多時候，多長的期間內再犯罪構成再犯，與學理或是價值判斷毫無關係，而只是單純因為實際開發、技術或是成本上的考量。例如，ORAS將再犯定義為於一年內再次被逮捕，就只是因為計畫已經預期在三年內完成系統的建置而已⁷⁰。

由前述可以知道，如何定義再犯，會直接影響到對於一個人風

⁶⁴ Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 76 (2017). 不過，作者同時提醒道，也有系統採取不同定義者。

⁶⁵ Latessa et al., *supra* note 53, at 15-16; COMPAS Risk & Need Assessment System: Selected Questions Posed by Inquiring Agencies, Northpointe, Inc. (2012), available at http://www.northpointeinc.com/files/downloads/FAQ_Document.pdf (last visited: 2019.12.30).

⁶⁶ Eaglin, *supra* note 64, at 76.

⁶⁷ *Id.* at 78.

⁶⁸ Latessa et al., *supra* note 53, at 16.

⁶⁹ COMPAS Risk & Need Assessment System: Selected Questions Posed by Inquiring Agencies, *supra* note 65.

⁷⁰ Latessa et al., *supra* note 53, at 15.



險評估的結果。舉例來說，如果再犯指的是被起訴，那麼某個曾經被逮捕，但沒有被檢察官所起訴的人可能就會被認為屬於低風險或沒有風險。然而，任何一種再犯的定義可能都有其不盡合理的地方。以現有風險評估系統較常使用的逮捕來說，其雖然有簡便及容易取得資料的優點，但是被拘捕並不當然就構成了「犯罪」，因為檢察官可能根本不會起訴該案件，起訴了，也可能獲致無罪判決，即使是被判決有罪，也可能在上訴後改諭知無罪判決，甚至是在有罪判決確定後，經再審，改判無罪。無論是那一種情形，該逮捕案件都不是刑法意義下的犯罪，是故，以逮捕來定義再犯罪，可能會有過度放大「犯罪」範圍的問題⁷¹。那是不是將所有逮捕、起訴、有罪判決等資料全部納入，就萬無一失，不會有漏網之魚呢？答案可能是否定的，因為這樣一來，就會有同一行為重複計算的問題。亦即，一個有罪判決確定的案件，可能同時歷經拘捕、起訴、有罪判決及確定等階段，事實上只有一個案件，但是在數據上會被計算為四次。此外，在什麼樣的期間內犯罪構成再犯，同樣會影響到再犯風險的評估。如果再犯指的是犯罪後一年內再犯罪，那麼犯罪後三年內有可能再犯之人，就不會是高風險的族群。

從前述可知，再犯的定義與再犯風險評估的結果密切相關，但僅有部分的系統對外說明了其決定再犯罪的概念或期間的原因。ORAS便解釋，認定只要是被逮捕就構成再犯罪的原因是，以起訴、有罪判決或是入監服刑等認定再犯罪，會需要更長時間蒐集及分析相關資料。另外，當特定人因為犯罪而被逮捕時，就已經嚴重

⁷¹ 從另一個角度來看，檢察官不起訴或是法院諭知無罪判決，並不當然代表沒有犯罪發生，可能只是因為證據不足或是其他程序上的事由（未經合法告訴或犯罪嫌疑人或被告已經死亡等）。

地影響到社會治安及秩序了⁷²。VRAG則解釋，其選擇以起訴作為認定再犯罪的原因在於，與有罪判決等其他因素相比較，起訴的相關資料較為正確⁷³。

(三)資料中影響再犯的因素為何

選擇及整合資料庫後，系統建置者接著必須要初步決定，在所選擇使用的資料中，哪一些是可能會影響到再犯率的事項。亦即，在什麼樣的情狀或是在什麼樣的條件下，特定人可能會再犯罪。如此一來，在建立預測模型（詳見後述）後，系統開發者就可以發現哪一些事項具有統計上的意義，進而確認真正會影響再犯的因素，用以在個案中預測特定人再犯罪的可能性⁷⁴。

風險影響因子的研究及選擇，並不是什麼新的事物。在過往，犯罪學家們就已經研究了各種可能影響再犯率高低的事項，如性別、種族、膚色、年齡、前科紀錄、犯罪類型、假釋撤銷紀錄、家庭、就業及交友等等⁷⁵。在這一個階段中被挑選者，才有機會被觀察到其對於再犯可能性的影響，未被選擇者，即使對於再犯可能性有所影響，也不會成為預測模型中所使用的風險因子。不過，在這一個階段再犯因子的選擇，並不是單純的科學或是數據上的判斷，而是由系統建置者依其個人的價值判斷或是主觀而決定。

⁷² Latessa et al., *supra* note 53, at 15-16.

⁷³ Eaglin, *supra* note 64, at 76.

⁷⁴ Barocas & Selbst, *supra* note 62, at 688.

⁷⁵ See, e.g., Nicholas Scurich, *Evidence-Based Sentencing: Public Openness and Opposition to Using Gender, Age, and Race as Risk Factors for Recidivism*, 40 LAW & HUM. BEHAV. 36, 37-38 (2016). 另外可以參考陳玉書，再犯特性與風險因子之研究：以成年假釋人為例，載：刑事政策與犯罪研究論文集，16期，頁3-5，2013年12月。

元照出版提供 請勿公開散布

(四)資料的取捨及修改

除了選擇風險因子外，由於資料中可能存在有錯誤，所以系統建置者還必須要修改，甚至是剔除某筆資料⁷⁶。舉例來說，一份未成年人，但曾經於一般監獄服刑過的紀錄，顯然就是不正確的。此時，建置者有兩個選擇，一是修改該筆資料，一是捨棄不使用之。如果是前者，必然涉及到系統建置者的主觀判斷⁷⁷。例如，將未成年人修改為成年人，或是將監獄修改為少年矯正學校。如果建置者選擇捨棄該筆資料不用，整體資料勢必會因而較不完整，缺少了數筆實際上曾發生的紀錄。無論建置者的決定如何，都會影響到風險評估系統最後運算結果的正確性。就此，不同的系統建置者可能有不同的決定，但是外界不當然會知道其在建置時，是選擇刪除或是如何修改錯誤的紀錄。

(五)建構預測模型

決定了資料庫，選擇了可能影響再犯率的風險因子後，開發者會建構一個統計模式，以確認所選擇的因子是否具有其統計上的顯著性（statistical significance），最後才確認系統中使用哪些風險因子。也因此，每一個系統的風險因子都不會完全相同。

從正面的來說，在選擇使用哪些風險因子時，多數的系統都納入了較多與犯罪紀錄相關的事項，如被拘捕、受有罪判決宣告及入監服刑等⁷⁸。以LSI-R為例，其便納入了被拘捕、起訴、受有罪判決、暴力行為紀錄及假釋條件違反等事項⁷⁹。COMPAS在預測審前

⁷⁶ Barocas & Selbst, *supra* note 62, at 684-85.

⁷⁷ Eaglin, *supra* note 64, at 80.

⁷⁸ Bernard E. Harcourt, *Risk as a Proxy for Race: The Dangers of Risk Assessment*, 27 FED. SENT'G REP. 237, 239 (2015).

⁷⁹ Hamilton, *supra* note 28, at 98.



保釋後，是否能到庭時，所考量的是起訴罪名、他案、其他被捕紀錄，以及是否曾不到庭等因素⁸⁰。這意味著，許多系統建置者認為，與犯罪相關的紀錄是重要或是具重要關聯性的風險因子。除了與犯罪紀錄相關的因素外，系統也會納入部分與犯罪不直接相關的事項。例如，VRAG納入了婚姻狀態、年齡、國小是否適應不良，以及十六歲前是否與生父母同住等⁸¹；COMPAS則使用了包括經濟狀況、就業、教育及家庭成員犯罪紀錄在內的15個因子⁸²。

從反面來說，系統建置者會排除掉在統計上不具顯著意義的風險因子。值得注意的是，部分因子，如種族或是性別，即使具有統計上的顯著性，但是因為憲法或是倫理的緣故，可能不會納入評估系統中⁸³。也就是說，某項因子與再犯等風險間可能有其統計上的關聯，但不當然就會使用於系統，還必須要顧及到統計以外的原因。至於要考量哪些因素，在政策上是個極為複雜且有不同看法的課題。舉例來說，美國聯邦量刑委員會挑選所採用的因子的標準，便相當嚴格，依其所制定的量刑準則，年齡⁸⁴、身心狀況⁸⁵、藥物或酒類濫用⁸⁶、在犯罪中的角色⁸⁷、犯罪紀錄⁸⁸，以及是否以犯罪

⁸⁰ Northpointe, Inc., *supra* note 39, at 27.

⁸¹ Melissa Hamilton, *Adventures in Risk: Predicting Violent and Sexual Recidivism in Sentencing Law*, 47 ARIZ. ST. L.J. 1, 14-15 (2015).

⁸² J. C. Oleson, *Risk in Sentencing: Constitutionally Suspect Variables and Evidence-Based Sentencing*, 64 SMU L. REV. 1329, 1399 (2011).

⁸³ *Id.* at 1352, 1380-82.

⁸⁴ U.S. Sentencing Guidelines § 5H1.1 (U.S. Sentencing Comm'n 2010).

⁸⁵ *Id.* § 5H1.3.

⁸⁶ *Id.* § 5H1.4.

⁸⁷ *Id.* § 5H1.7.

⁸⁸ *Id.* § 5H1.8.

維生等事項⁸⁹，可以作為量刑參考之用，被告的教育、專業⁹⁰及就業情形⁹¹，可以作為決定釋放條件時的參考，家庭關係，原則上不得作為量刑的因素⁹²，至於種族、性別、國籍、宗教信仰及社經地位則絕對不得於量刑時納入考量⁹³。相比較下可以發現，美國聯邦量刑準則排除許多其他評估系統所會涵蓋，且與再犯或是暴力犯罪相關的風險因素⁹⁴。

⁸⁹ *Id.* § 5H1.9.

⁹⁰ *Id.* § 5H1.2.

⁹¹ *Id.* § 5H1.5.

⁹² *Id.* § 5H1.6.

⁹³ *Id.* § 5H1.10.

⁹⁴ See Monahan, *supra* note 31, at 397-98. 我國雖然沒有量刑準則的規範，但司法院於2011年開始，建置了「量刑趨勢建議系統」，供法官量刑時參考。系統以統計回歸方法分析實務判決，參考了政府機關代表、學者及民間團體代表等焦點團體的意見後，針對各個罪名，選擇了不同的量刑因子，並調整其影響力大小。以刑法第221條的強制性交罪為例，系統就刑法第221條所選定的參考因子有行為人行為時是否為十八歲以上未滿二十歲、妨害性自主前案紀錄次數（一次、二次或三次以上）、被害人數（一人、二人或三人以上）、是否坦承犯罪事實且始終一致（偵查中坦承，或審判中始坦承），以及與被害人之間的關係（有監督保護關係、有家庭成員關係、網友，或其他於日常生活中的信賴關係）等事項；針對刑法第57條，系統則考量了是否以身體部位以外之部位進入、是否在一行為中反覆實施性交（犯罪行為態樣）、是否造成嚴重身體傷害、是否造成嚴重精神創傷、是否仍有健全成長可能（犯罪行為造成之損害）、在偵審過程中造成被害人二度傷害、與被害人達成和解、道歉且得被害人宥恕（犯後態度）等因子。司法院，量刑趨勢建議系統（刑法第221條第1項妨害性自主罪），http://sen.judicial.gov.tw/pub_platform/sugg/221I.html，最後瀏覽日：2020年1月12日。此外，司法院同時製作了「刑事案件量刑審酌事項參考手冊」。兩者間不同的是手冊主要是根據刑法第57條，列舉量刑時應參酌的具體事項。以毒品犯罪為例，手冊上所建議的考量事項就包括了是否是為了控制他人、是否係為了賺錢購買毒品供自己施用而販賣（犯罪之動機、目的）、販賣數量、賺取利潤、是否從境外取得毒品、販賣的對象、是否利用網路販賣（犯罪之手段）、是否以販毒為



由前述的討論可以知道，評估系統中要採納哪一些因子來計算風險值，並非完全出於關聯性或統計上有無顯著意義的考量，其中還會涉及到政策、規範及價值性的判斷。類似的考量，也已經在個案產生爭議。舉例來說，在State v. Loomis案中，由於威斯康辛州所使用的風險評估系統COMPAS，將性別作為風險評估的事項，而下級審法院在量刑時又參酌了COMPAS的風險評估，被告於是爭執，法院的量刑方式違反了正當法律程序的要求，侵害了被告不

業、是否同時有其他財產或暴力犯罪（犯罪行為人之生活狀況）、有無轉讓或販賣毒品的前科紀錄（犯罪行為人之品行）、是否具有化工或醫藥等毒品相關知識（犯罪行為人之智識程度）、買受人施用後是否死傷（犯罪所生之危險或損害），以及是否供出毒品來源及是否自白（犯罪後之態度）等。參司法院，刑事案件量刑審酌事項參考手冊，頁1-2，2018年8月。雖然量刑趨勢建議系統或是刑事案件量刑審酌事項參考手冊，都沒有使用到演算法或是人工智慧，但是其對於風險或是量刑事項的選定，都還是有其類似的部分。亦即，無論是建議系統或參考手冊，雖然使用了統計方法（如回歸分析等），但是其中的量刑因素都是由人所決定，因而帶有一定程度的主觀取捨。是故，雖然刑度的建議是以系統或是手冊的方式呈現，但仍非完全的客觀或中立，而是帶有個人價值上的判斷。另外，需要留心的是，最高法院認為，量刑趨勢建議系統的運作結果，僅供法官參考，並沒有拘束效力。法院的量刑即使不同於量刑趨勢建議系統所建議的刑度，也不當然違法，不屬得上訴第三審的事由。最高法院108年度第3728號判決參照。另外，司法院也建置了「量刑資訊系統」。這一個系統針對數個罪名，統計了不同的量刑因子對於刑度的影響。以刑法第221條的強制性交罪為例，便納入了累犯、公務員（加重事由）、未滿十八歲之人、滿八十歲之人、中止犯、自首（減輕事由）、強制力態樣、犯罪態樣（性器進入性器，或性器進入肛門等）、前案紀錄、所生損害（身體傷害、懷孕等）及犯後態度。選擇因子後，系統便自動產出類似案件的數量及量刑分布情形。司法院量刑資訊服務平臺，<https://sen.judicial.gov.tw/>，最後瀏覽日：2021年7月24日。與量刑趨勢建議系統相同的是，最高法院同樣認為，量刑資訊系統只是供法官參考，並不限制法官量刑上的裁量權限，所以當事人不能以法院量刑與系統產出結果不同，指摘量刑違法，據以提起上訴。最高法院110年度臺上字第469號判決參照。

因其性別而受有不同刑度處罰的權利⁹⁵。檢察官則回應道，由於男性及女性有著截然不同的再犯率，復歸及矯治可能也各異，為要促進風險評估的正確性，必須要考量受判決人的性別，否則預測結果就會有很大的誤差⁹⁶。

針對這個爭點，當事人雙方都不爭執的是，統計上男性有較高的再犯率，且更容易犯下暴力犯罪。學界研究的結果，也有著相同的結論⁹⁷。不過，被告主張，即使如此，以性別作為評估風險的事項，仍違反了正當法律程序的要求⁹⁸。針對這一個爭點，威斯康辛州最高法院的看法是，COMPAS將性別納入風險評估的事項，有其事實上的依據。研究及事實都顯示，在評估風險時，若完全不考量性別，無論是在男性或女性，都會大幅地提高誤判的機率。由於將性別加入風險評估系統中可以大幅地增加其正確性，所以性別的考量是為了制度及被告的利益，而非歧視，所以不違反正當法律程序原則⁹⁹。Loomis案是州法院的判決，但卻極具指標性意義，因為其是在近年實務上大量使用人工智慧進行風險評估後，首次法院針對其中爭議，作成判決，所以相當具有指標意義。另外，雖然威斯康辛州最高法院判定風險評估系統以性別作為風險因子不違反正當法律程序，但是可以預見的是，這一個案件只是開始，學者及實務就此一定會有不同的看法，類似的爭議在個案中還是會不斷地出現。

⁹⁵ *Loomis*, 371 Wis.2d at 267.

⁹⁶ *Id.* at 267-68.

⁹⁷ *Id.* at 268.

⁹⁸ *Id.* at 269. 值得注意的是，被告並未主張納入性別為風險評估因子違反了聯邦憲法對於平等原則的要求。

⁹⁹ *Id.* at 270. 除此之外，法院也指出，被告並未盡到舉證責任，證明法院以性別作為量刑的因素。事實上，法院判定被告有再犯的高度風險的因素是犯罪類型、犯罪紀錄、管束紀錄（history of supervision）、本案與犯罪紀錄間的關聯，以及評估系統的結果等。See *id.* at 271.



(六)風險因子是否及如何加權

在決定於風險評估系統使用哪些風險因子後，系統建置者接著要決定給予因子什麼樣的權重。就此，主要有兩種模式。一是Burgess法，一是加權Burgess法¹⁰⁰。前者給予每一個風險因子相同的分數；後者，望文生義，就是給予部分的因子較高的數值。舉例來說，某一個風險評估系統使用了前案紀錄、就業、就學、教育程度、家庭狀況、婚姻、固定住居所及經濟狀況等因子，依Burgess法，那麼每一個因子都會固定是一分。如已婚，零分；未婚，則是一分。依加權Burgess法，開發者就會給予某些因子較高的分數，如有前案紀錄者，兩分；沒有者，零分，但其他的因子都是一分等。多數的風險評估系統，都是使用Burgess法，因為其較為簡便直觀¹⁰¹。

選擇使用Burgess法或加權Burgess法本身是系統建置者的主觀上的決定，會直接影響到某些風險因子對於風險評估運算的結果。如果建置者選擇了加權Burgess法，那麼他就還要決定，要加重哪些風險因子的權重，以及如何加重。這些決定都涉及到建置者的價值判斷，不是純粹客觀的技術或是科學數據。

(七)個案數據的取得及輸入

無論系統建置者選擇了什麼樣的風險因子及是否予以權重，都

¹⁰⁰ James L. Johnson et al., *The Construction and Validation of the Federal Post Conviction Risk Assessment*, 75-SEP FED. PROBATION 16, 17 (2011).

¹⁰¹ E.g., Pa. Comm'n on Sentencing, Risk/Needs Assessment Project: Interim Report 4: Development of Risk Assessment Scale 8 (2012), available at <http://pcs.la.psu.edu/publications-and-research/research-and-evaluation-reports/risk-assessment/phase-i-reports/interim-report-4-development-of-risk-assessment-scale/view> (last visited: 2020.01.14).

元照出版提供 請勿公開散布

必須要有人取得及輸入個別風險因子的相關資訊，才能透過其演算法，計算個案的風險值。如何取得及輸入，不同的評估系統有著不同的作法。部分要求必須進行評估前晤談（interview），以蒐集相關資訊，輸入於風險評估系統中，如印地安納州（Indiana）¹⁰²。也有由受評估者自行填寫表格或問卷，如COMPAS¹⁰³。部分則是輸入已經存在的資料，如前科紀錄或到場情形等，不另外蒐集額外的資訊或是與相對人進行晤談，如Laura and John Arnold基金會所開發的系統¹⁰⁴。由於實際的應用上，有著不同的資料或數據的輸入方式，風險評估的結果可能就會有所不同。

詳細地來說，部分評估系統之所以直接輸入既有資料，而不是另外進行晤談，並不是因為前者是較好或更為準確的方式，而僅僅是因為成本的考量¹⁰⁵。但是，如果晤談或是郵箱對人填寫表格或問卷等方式是更好的蒐集資訊方式，但卻因為成本考量而改用現有資料，就必然地犧牲了系統運作結果的準確性，進而左右了犯罪嫌疑人或被告是否會受到拘禁或是釋放條件，影響其人身自由等權益。

再者，使用晤談或是由受評估人填寫表格的方式來蒐集所要輸入系統的資料，也會受到人的主觀因素的影響。以晤談來說，實施晤談人員是否受有嚴格、完整且有系統的培訓，發問方式及態度等

¹⁰² Eaglin, *supra* note 64, at 85.

¹⁰³ COMPAS Risk & Need Assessment System: Selected Questions Posed by Inquiring Agencies, *supra* note 65, available at http://www.northpointeinc.com/files/downloads/FAQ_Document.pdf (last visited: 2020.02.06). 雖然COMPAS有提供晤談的表格，但晤談並不是絕對必要的程序，評估機關仍可視情形決定是否進行晤談。

¹⁰⁴ See Laurn P. Gouldin, *Defining Flight Risk*, 85 U. CHI. L. REV. 677, 718 (2018).

¹⁰⁵ See *id.*

等，會左右了晤談的結果。除此之外，如何理解受評估人的回答及反應，怎麼樣將之轉化問題的答案及風險因子的數值，也會影響到輸入評估系統的資訊是否正確及完整，以及系統對於特定人的評估結果。同樣地，表格的設計或詢問的方式、受測者對於問題的理解及答案的表達方式，甚至是資料輸入者對於答案的理解，都會決定了系統所運算出來的風險值高低。

(八)建構風險評估工具

確認評估系統所要使用的風險因子及其權重後，系統建置者接著會以所選擇的因子及演算法（algorithm）建構一評估工具，計算個別案件中特定人的風險值。這一個步驟中，有著幾點值得留心的地方。首先，絕大多數的系統開發及設計者不會對外揭露所使用的演算法或人工智慧，所以外界通常不會知道風險評估系統使用了什麼樣的程式或編碼。第二，風險評估工具並不是真的去計算個案中的特定人的分數，而是透過演算法及所選擇的風險因子，產出與特定人具有類似條件或情形者，有多高的風險值。亦即，由於受評估人與這一個族群有相仿的背景，所以有著相同的風險值。這裡的風險（值），依照系統建置目的及功能的不同，可能是指無正當理由不到庭、逃亡、違反釋放條件或是再犯的可能性高低等等。

(九)風險群組的建立

風險評估系統的運算結果，並不是單純的數值，也會涉及到人的主觀決定及價值取捨。詳細地來說，系統針對個案計算出數值後，為了後續應用上的方便起見，多會再依其數字的高低，將個案分為不同程度的群組，如高、中及低度風險¹⁰⁶。至於什麼樣的風

¹⁰⁶ See Pa. Comm'n on Sentencing, Risk/Needs Assessment Project: Interim Report 8: Communicating Risk at Sentencing 4 (2014), available at <http://pcs.la.psu.edu/>



元照出版提供 請勿公開散布

險值屬於那一個群組，與電腦程式、科學或數學無關，純粹是人的判斷，取決於人們（或者是系統建置者）對於風險的容忍程度。舉例來說，當人們覺得20%的再犯可能性已經無法忍受的風險時，就會認為其係高風險；相對地，如果覺得20%的機率並不高，就會認為其只是中度，甚至是低度風險。至於為什麼高度風險是30%以上？為什麼不是40%？為什麼10%的再犯可能性是低度風險，11%就是中度風險？多低的犯罪可能性是可以忍受的風險？多高又是人們所不能容忍的可能性？這些都涉及到人的價值判斷，以及整體社會所能夠容忍的風險高低，都不是純粹的數學計算或是電腦運算的結果。事實上，不同的社會、文化或族群，在不同的時代，所會提出的答案都不盡相同。然而，被歸類為不同風險群組等於是決定了受評估人的風險評估結果，進而可能影響到其人身自由是否會受有拘禁，以及若是被釋放時所必須要遵循的條件。

綜上可知，風險評估系統的運算結果看似客觀，是以電腦程式，依據風險因子，將特定人分類為不同的風險群組。然而，風險群組的界定（分），本質上是系統建置者的價值判斷及主觀決定，但卻又直接影響到風險評估的結果，甚至是受評估者的人身自由。

(十)正確性（accuracy）的審查

風險評估系統的建置及應用，應包含有評估結果正確性的審查。其中的原因並不難理解，因為系統評估結果若有誤，可能會縱放了可能再犯罪之人，危害社會大眾，更嚴重的是，會使得低風險，甚至是沒有風險之人，身陷囹圄，不見天日。在進一步討論風險評估系統的正確性的審查前，有必要先瞭解評估結果正確或是錯

publications-and-research/research-and-evaluation-reports/risk-assessment/phase-i-reports/interim-report-8-communicating-risk-at-sentencing/view [hereinafter INTERIM REPORT 8]. (last visited: 2020.02.07).

元照出版提供 請勿公開散布

誤的意涵為何。亦即，在什麼情形下，我們可以說評估的結果正確（或錯誤）？

評估結果的錯誤，會有兩種類型，一是偽陰性（false negative），一是偽陽性（false positive）。前者指的是誤高風險者為低風險，結果就是予以釋放，或是附較為寬鬆的釋放條件。這一種錯誤所產生的成本，會是由被害人或是整體社會所承擔。被害人可能因而受有身體、財產、自由、心理或甚至是生命上的危害，犯罪活動同時也會破壞了社會整體的秩序或安全。後者則是誤將低風險者判定為高風險，會拘禁無或是低度再犯可能者，或是不予釋放。此時，評估錯誤的成本是由受評估者所承擔，其可能會因而失去了人身自由，或是受到各樣的不利益或限制（必須要定期報到、接受治療或是在特定時間內不得離開住所等），執法機關或法院會因而耗費多餘的成本拘禁被告或確保其遵守釋放條件。無論是何種錯誤，都會大大地減弱了一般人民對於刑事司法制度的信心¹⁰⁷。

是不是評估系統的結果有誤時，都能夠被發現？答案可能是否定的。以偽陽性來說，由於被誤判為具有高風險的被告或犯罪嫌疑人，都已經因為預測的結果受有拘禁，所以不可能再犯罪，也不會有違犯釋放條件可言，所以無從判斷系統的預測是否正確。但這是否意味著，我們至少可以從再犯罪來發現偽陰性的錯誤，或是審查評估系統這部分的正確性？針對這一個問題，答案可能還是否定的。首先，被告或犯罪嫌疑人的再犯罪，要如何認定？亦即，被評估為低風險而釋放之人，可能因為涉嫌不法活動而被拘捕、起訴、被判決有罪或是有罪判決確定，要以那一個程序中的決定認為是再犯罪？¹⁰⁸ 無論是選擇那一個，都會屬於政策目的的選擇、價值判

¹⁰⁷ 然而，相較於偽陰性，偽陽性的錯誤在制度上幾乎無法確認或統計。

¹⁰⁸ Eaglin, *supra* note 64, at 91.



斷，以及社會對於風險所能承受的程度。此外，犯罪可能會因為種種原因而未被國家機關所察覺，如被害人未報案，或是被害人不知道行為人是誰。在這些情形下，風險評估系統所產生的偽陰性的錯誤，也不會被發現¹⁰⁹。

風險評估系統的結果是否有誤，還會涉及到刑事程序上對於「正確」的理解，這至少包含了兩個面向。一是越能夠避免有罪之人逍遙法外，就越是正確的刑事司法。也就是說，正確的刑事程序應該是要竭力使罪犯伏法；另一種觀點則是，正確意指使無辜之人不會身陷囹圄。正確的刑事司法，應該要能夠還給無罪之人一個清白，不會屈枉冤抑。若將正確定義為前者，就會「寧可錯殺一百，不可錯放其一」；如果是後者，則是「寧可錯放一百，不可錯殺其一」¹¹⁰。從這一個角度來看，評估系統的結果是否正確（錯誤），也必然會受到社會大眾或是理論上對於刑事程序的基本價值的設定所影響。

此外，即使風險評估系統的結果「正確」，也無法說明風險群組的分類。亦即，究竟什麼樣的風險值屬於高、中或低風險族群，與評估系統「正確」間並沒有直接的關係。這是因為，什麼樣的犯罪嫌疑人、被告或受判決人屬於那個風險族群，是價值判斷及主觀選擇，與評估系統對於風險值的計算結果是否正確間，沒有其關聯性。也因此，從事後的結果，很難就據以認定系統運作結果是否有誤。

在實際的應用上，風險評估系統的正確性也已經成為了個案中的爭點。在Loomis案中，被告提出了一份加州的研究報告，該份報

¹⁰⁹ Christopher Slobogin, *Dangerousness and Expertise Redux*, 56 EMORY L.J. 275, 291 (2006).

¹¹⁰ 王兆鵬等，刑事訴訟法（上），頁4-5，2020年3月，5版。



告雖然肯定COMPAS是可靠的工具，其再犯風險的預測達到了傳統所要求的0.7的標準，但是其對於暴力風險的預測並沒有這樣高的正確性。此外，報告作者也指出，沒有任何的證據可以證明，不同的施測者間能夠以一致的方式使用COMPAS。不過，數個在實務上應用COMPAS的州，都針對了該系統進行了效度（validation）評估，結果顯示COMPAS是相當正確的風險評估工具。紐約州（New York）的研究報告更是指出，無論是在再犯評估的有效性與預測的正確性，都有令人滿意的表現¹¹¹。從實務上的這一些爭議及評估的結果可以知道，風險評估系統是否正確或能否有效地評估特定人的再犯率及風險高低，會是實際應用上的問題及爭議所在。

肆、運作上不透明／黑盒的疑慮

從前述的分析及討論可以知道，風險評估系統在建置的過程中，在每一個重要的階段幾乎都無可避免地會受到人的主觀因素的左右，所以並不像其外觀看來地客觀或毫無偏見。除了在建置過程中會受到參與在其中的人員的價值判斷影響外，由於所使用的人工智慧技術的緣故，外界難以知悉系統產出結果的原因，這也就產生了風險評估系統運作上的不透明或是黑箱（盒）（blackbox）的現象，進而有著使用系統的正當性的擔憂。以下擬先更進一步地解釋人工智慧所會造成的不透明或黑盒現象，其所引發的種種疑慮，再進一步提出可能的解決途徑。

¹¹¹ Loomis, 371 Wis.2d at 262.

元照出版提供 請勿公開散布

一、現象及其成因

利用人工智慧所建置的風險評估系統，會有著系統運作上不透明或是黑盒運作的疑慮。這是因為，許多人工智慧都使用了神經網路（**neural network**，又稱人工神經網路或類神經網路）的技術來進行運算。神經網路的運作原理類似於人的大腦。一般軟體是就所輸入的數值，依預先所編寫的程式碼，產出運算的結果。輸入的資料及運算結果，保持不變。例如，輸入2加上2，答案永遠會是4。但利用了神經網路的人工智慧則不是如此。神經網路模仿了人大腦的神經系統，其中有輸入層、數量不一的隱藏層（通常在兩層以上）及輸出層。每一層中都有大量的神經元，神經元之間彼此連結，並且透過運算的結果（正確或是錯誤等）等外在資訊，不斷地調整隱藏層中神經元連結的權重等內部的結構，以提高運算的結果的正確率。這個過程，不需要人為的介入，是人工智慧自己的學習（訓練）。一個學習好的人工智慧，可以有很好的判斷或預測結果，但是由於其中有著大量的神經元及連結，以及因為不同的運算結果而反向調整權重的情況，所以從外部很難（甚至是不可能）知道為什麼人工智慧可以達到這樣的運算或是預測正確的效果¹¹²。這就像是，絕大多數的小孩子都知道不能摸火，或是饅頭可以吃，但是如何發展出這一個認知的過程（神經元連結），每一個人可能都不一樣。以不能摸火的例子來說，有人可能是因為爸爸媽媽說不可以，所以知道不能碰火，有人可能是因為想去摸，被處罰了，有人可能是因為看到別人摸了火後的反應，有人可能是因為自己去摸了，不一而足。外界難以得知運算結果的原因及過程，是人工智慧的本質所致，也因此，當國家機關在決定（策）過程中使用人工智慧時，

¹¹² Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829, 1832 (2019).

元照出版提供 請勿公開散布。
會有著「黑盒疑慮」(the black box problem)¹¹³。

二、相關疑慮及知悉原因的需要

由於人工智慧有著黑盒運作的問題，以人工智慧所建置的風險評估系統，也就無可避免地受到外界的質疑。基於下列實務及學理上的原因，風險評估系統的導入或使用應有其透明性(transparency)的要求。

(一)錯誤或是不合理之處的即時發現

外表看來，風險評估系統的運作客觀公正，沒有人為主觀因素介入，但是系統並不會(無法)交待為什麼會得出某個風險數值或特定人屬於某個風險族群的理由，所以可能就會有「系統是否如其所說明或設計的方式運作」的疑慮。此時，如果系統中存有隱而未現，不適當的主觀因素或是錯誤，因而可能造成不當且不合法的歧視或偏見，就會難以即時發現¹¹⁴。

再者，無法知悉系統運算結果的原因，會使得人們無從判斷結果是否合理，以及系統是否需要調整。這是因為，很多時候，結論正確固然重要，但更重要的是，得到結論的原因及過程為何，因為如此，我們才能知道結論是否合理，以及其是否值得參考。舉例來說，在判斷圖片中的動物是貓或是老虎時，電腦可能是因為圖片中出現了老鼠或是罐頭，而認為牠是貓。如果是基於這樣的原因辨識出圖片中的動物是貓，我們也不會認為該程式可以應用在其他物件的辨識上，因為圖片中若是沒有老鼠或罐頭，程式可能就會顯示該

¹¹³ Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1254 (2008).

¹¹⁴ See Omer Tene & Jules Polonetsky, *Judged by the Tin Man: Individual Rights in the Age of Big Data*, 11 J. TELECOMM. & HIGH TECH. L. 351, 358 (2013).



動物不是貓。同樣地，我們不只需要知道特定人屬於那一個風險群組，還需要知道為什麼，如此一來，才能夠確保風險評估系統所審酌的因素及運算的過程合理，不會基於全然無關（如所駕駛的汽車的顏色）或是不妥適的原因（如政治傾向）¹¹⁵。特別是風險評估系統的運作結果，可能會影響到受評估者的自由或財產等權益時，我們更會需要知道系統形成結果的原因為何。從另一個角度來說，知道風險評估系統（或人工智慧）的運算結果的原因，也可以幫助外界發現事物中以前所未察覺到的現象或規律¹¹⁶。亦即，由於人工智慧可以從大量的過往資料，找到其中的規律及模型，據以預測現有的事物可能的後果，所以其可能因而可以幫助人們認識到嶄新的知識或事物間不曾被觀察到的關聯。

（二）透明性及訴訟上的防禦權

黑盒運作的問題，連帶地還會影響到後續的救濟。當風險評估系統的運作是個外界無法得知其中運作過程的黑盒子時，即使容許相對人對於運作的結果聲明不服，提起救濟，程序上也很難有實質意義的審查或是監督，因為沒有人工智慧或是電腦程式相關專業知

¹¹⁵ 當無從得知人工智慧產出的結果的原因時，會很難取信於外界，特別是當政府機關在決定過程中使用此一工具時。無法（完整）說明決定的理由，會使人們覺得決定恣意，不公平及不合理。也因此，對於可解釋的人工智慧（explainable AI, xAI）的呼聲，也就應運而起。科技部提出的「人工智慧科研發展指引」（AI Technology R&D guidelines）中，有著「可解釋性」（explainability）的要求〔科技部，人工智慧科研發展指引（2019年9月版），<https://www.most.gov.tw/most/attachments/53491881-cb0d-443f-9169-1f434f7d33c7>，最後瀏覽日：2019年12月2日〕。

¹¹⁶ Finale Doshi-Velez & Been Kim, Towards a Rigorous Science of Interpretable Machine Learning 1, 3 (2017), available at <https://arxiv.org/pdf/1702.08608.pdf> (last visited: 2020.06.08).



識之人，幾乎不知道為什麼系統會產出那樣的結果，或是為什麼風險評估系統判定特定人為高、中或低風險¹¹⁷。是故，若是在刑事司法程序中導入風險評估系統的應用，掌握系統運作結果的原因才能讓外界檢視結果所依據的事實基礎為何，以及其決定是否合理，並使法院瞭解何以系統會產出該結果，被告等聲明不服之人及其律師也才能夠有效地提出攻擊或防禦，指出系統中可能有誤之處¹¹⁸。對法官等官員來說，所使用的工具的可解釋性，有其重要性¹¹⁹。

使用風險評估系統的透明性，不只是學理上的討論或倡議，實務上也已經注意到其重要性，甚至成為具體個案中的爭議所在。Loomis案中的一個爭點便是與系統的透明性有關。在這一個案件裡，被告主張，COMPAS的結果是量刑前調查報告（PSI）的一部分，但是他沒有任何的資料可以用來檢視COMPAS的正確性，所

¹¹⁷ Jason Tashea, Courts Are Using AI to Sentence Criminals. That Must Stop Now, *Wire* (Apr. 17, 2017), available at <https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/> (last visited: 2019.09.16).

¹¹⁸ Deeks, *supra* note 112, at 1845-46. See, e.g., Robin A. Smith, Opening the Lid on Criminal Sentencing Software, *Duke Today* (July 19, 2017), available at <https://today.duke.edu/2017/07/opening-lid-criminal-sentencing-software> (last visited: 2019.12.02).

¹¹⁹ 不過，以現有的救濟機制來說，即使要求使用風險評估系統的國家機關必須要說明相關事項，並讓允許受評估人聲明不服，其效果可能還是相當有限。這是因為，法院可能沒有審查風險評估系統的運作及為什麼產出該結果，因而只能作形式審查（是否輸入正確的數據、是否按照相關規則進行評估等）。如此一來，救濟機制就只是聊備一格，可有可無的裝飾品。就當事人來說，即使系統建置者同意提出系統的程式編碼、所使用的風險因子及權重等資料，供當事人及救濟機關檢視，當事人勢必要花費極高的代價，借助專家學者，才能夠檢視所使用的風險評估系統中有無錯誤或是不當之處。如此一來，能否有意義地使用救濟程序，聲明不服，等於是取決於相對人的資力高低，是否合乎平等原則，不無疑慮。See, e.g., Dana A. Remus, *The Uncertain Promise of Predictive Coding*, 99 IOWA L. REV. 1691, 1705 (2014).

以法院所進行的量刑程序侵害了其受以正確資訊受有刑度量定的權利¹²⁰。就此，Northpointe（COMPAS的所屬公司）的解釋是，COMPAS是封閉性工具（proprietary instrument），其相關資訊屬於營業秘密（trade secret），所以無法向被告揭露各個事由的權重或是程式如何計算風險值¹²¹。威斯康辛州最高法院判定，在本案中，被告雖然無法檢視及爭執COMPAS如何計算風險值，但是他仍然可以知悉量刑前調查報告中關於風險值的記載，並就其表示意見或不服。法院援引聯邦最高法院的判決指出，只有在被告完全沒有機會針對法院量刑所依據的資訊聲明異議、解釋或說明時，才會違反了正當法律程序的要求¹²²。但在本案中，Northpointe公司所出版的二〇一五年COMPAS應用指導已經說明了風險值主要是依據犯罪紀錄，並參酌部分的動態參數（如共犯或藥物濫用等）等事由¹²³。這些事由都是經由被告所填答的問卷所知悉，所以被告並不是沒有機會檢視程式所使用的資料，並就風險值的正確與否表示不服。再者，被告完全可以檢視法院所使用的量刑前調查報告，並爭執其正確性¹²⁴。基於上述的理由，威斯康辛州最高法院並未接受被告的這點主張。但是，從法院的說理可以知道，在量刑等刑事程序中使用風險評估系統之類的工具時，透明或外界得以檢視會是程序上一個重要的要求。

針對這一個議題，美國聯邦政府也承認，風險評估系統的不透

¹²⁰ *Loomis*, 371 Wis.2d at 257-59.

¹²¹ *Id.* at 258.

¹²² *Id.* at 258-59 (citing *Gardner v. California*, 430 U.S. 349, 362 (1977)).

¹²³ *Id.* at 259.

¹²⁴ *Id.* at 260-61. 是故，法院同時也說明道，由於量刑前調查報告是量刑的主要依據，所以如果未向被告揭露量刑前調查報告的內容，其程序便有所違誤，悖離了正當法律程序的要求（*Id.* at 257-58）。

明，使得被告無從知悉輸入程式的數據或是計算的公（程）式時，就會有違反正當法律程序的疑慮¹²⁵。從實務運作來說，著眼於外界難以檢視人工智慧內部事項及運作結果，美國許多的司法單位已經改使用更為透明，向外界公布所採納的風險因子的評估工具，如 **Public Safety Assessment**¹²⁶。

除了前述政策面的理由外，也有學者認為，從保障被告對質詰問權的角度，同樣可以推導出對於風險評估系統的透明性的要求。詳細地來說，針對人證，刑事程序中有諸多機制，來確保其內容的可信，如交互詰問、具結及傳聞法則等等。但是，顯而易見的是，如果在刑事程序中導入風險評估系統等電腦程式，前述的機制對於確認系統運作結果的可信性，幾乎沒有任何的作用可言¹²⁷。也因此，美國有學者主張，聯邦憲法第6增修條文中的對質詰問條款，不只是在保障人民對質及詰問權，更是使其得以詳盡地審視所有不利於己的證據¹²⁸。因此，當決定在刑事程序中使用人工智慧所建置的風險評估系統時，依對質詰問條款，應公開程式的原始碼及其他相關事項。如此一來，受評估人才能夠實質地接觸到國家所使用的程式，並得以審視及挑戰運算結果的有效性及可信性¹²⁹。再

¹²⁵ Brief for the United States as Amicus Curiae at 18, *Loomis v. Wisconsin*, 137 S. Ct. 2290 (2017) (No. 16-6387), 2017 WL 2333897.

¹²⁶ Creating a Fairer Pretrial System, Arnold Ventures (Dec. 1, 2017), *available at* <https://www.arnoldventures.org/stories/creating-a-fairer-pretrial-system> (last visited: 2019.12.04); Public Safety Assessment: Risk Factors and Formula, Pub. Safety Assessment, *available at* <https://www.psapretrial.org/about/factors> [<https://perma.cc/W3P3-4UJN>] (last visited: 2019.12.04).

¹²⁷ See Brian Sites, *Rise of the Machines: Machine-Generated Data and the Confrontation Clause*, 16 COLUM. SCI. & TECH. L. REV. 36, 42 (2014).

¹²⁸ See David Alan Sklansky, *Hearsay's Last Hurrah*, 2009 SUP. CT. REV. 1, 71 (2009).

¹²⁹ See Edward K. Cheng & G. Alexander, *Beyond the Witness Bringing a Process*

者，公開原始碼及相關事項，也較有機會發現系統中可能的錯誤、偏見或是不當之處¹³⁰。不過，法院面對類似主張的回應多是，電腦程式不屬於傳聞，或是系統的運算結果不構成對質詰問條款的證述（testimonial statement），所以沒有傳聞法則或是聯邦憲法第6增修條文的適用¹³¹。

如同美國法院的判決，就電腦程式或是風險評估系統適用對質詰問條款等規範的主張，可能根本不合於該條款的要件（如「證述」等），過於速斷，但是值得留心的是，之所有會有這樣的主張，是因為在政府機關使用風險評估之類的系統時，被告很難有意義或實質地檢視其中是否存在有錯誤或是不適當的地方。即使法院傳喚操作系統的人員到場、具結及接受交互詰問，其也只是陳述其是否具有操作的資格、以什麼樣的方式操作，以及是否正確地輸入指令及數據，徒具形式意義，因為系統運作結果是否可靠，主要是在系統本身，操作的人員的影響不大。是故，若是在刑事司法程序中導入風險評估系統，協助相關官員作成決定，法制面上就應該要確保被告、處分相對人或是外界能檢視系統的可信性及正確與否，以確保程序的正當及有效保護人民的權益¹³²。

伍、風險評估系統在使用上應依循的原則

由於人工智慧等電腦技術的進步，司法官員已經能夠在刑事程

Perspective to Modern Evidence Law, 97 TEX. L. REV. 1077, 1107-08 (2019).

¹³⁰ See, e.g., Erin Murphy, *The Mismatch Between Twenty-First-Century Forensic Evidence and Our Antiquated Criminal Justice System*, 87 S. CAL. L. REV. 633, 661 (2014).

¹³¹ See, e.g., *People v. Goldsmith*, 326 P.3d 239, 249-50 (Cal. 2014); *State v. Melsky*, 2013 WL 1776037, at *4 (N.J. Super. Ct. App. Div. Apr. 26, 2013).

¹³² See Andrea Roth, *Trial by Machine*, 104 GEO. L.J. 1245, 1301 (2016).



序中應用風險評估系統。這些系統除了在建置過程中可能會因為人的價值抉擇而帶有偏見，以及運作時會有不透明的問題外，還會有以系統運算結果作為決定因素、系統功能與程序目的間的不一致，以及過度依賴系統等爭議。也因此，在刑事程序中使用COMPAS之類的風險評估程式時，也應該要注意到這些問題，遵循一定的原則及程序要求，以避免不當侵害人民人身自由等權益。

一、透明性的要求

如同先前所討論地，由於人工智慧的技術的緣故，使用其所建置的風險評估系統會有不透明或黑箱的現象，因而會有不容易判斷其運算結果是否正確、合理，以及使得後續的救濟難以發揮應有的效果等問題。是故，在實務運作上使用風險評估系統時，應對外揭露必要的相關事項。

(一)相關資訊及事項的公開

在刑事程序中使用風險評估系統時，系統的相關事項，對外應具其透明性，如此一來，社會大眾對系統才能有一定程度的瞭解，並對其表示意見，並確保評估系統的使用不會與多數人的價值觀念相衝突，並使被告或其他受測者於事後得提起有實質意義的救濟。是故，諸如所使用的資料庫、所預測的風險、風險的定義、風險因子、輸入系統的個案數據如何取得（訪談、問卷或是使用既有資料）、風險群組的界分及結果的事後審查機制等資料，都應讓外界得以知悉。

(二)應對外揭露採用風險評估系統的標準及程序

除了在建置的過程外，採用風險評估系統的標準、條件或是程序，也應有透明性的要求。以美國目前的實務為例，多數州立法者



並未決定所使用的風險評估系統應具備的條件或是選擇時應遵循的程序，結果就會是由其他機關，如法院或矯治單位自行決定要使用的系統，例如，俄亥俄州就是由矯正機關自行選擇¹³³。不過，這樣一來，外界便無從得知決策的原因，更遑論參與其中。為了使風險評估系統的採用更為透明，在選擇所使用的系統時，應使社會大眾有表示意見的機會，如此一來，公眾的意見、想法及價值觀才能夠更為完整地反應在系統及其應用上¹³⁴。因此，當國家機關在刑事程序中使用風險評估系統時，必須要知道系統得到運算結果的原因，以確保其可信性及公平，不能只是知道建議或是數據而已。同樣地，被告也才能夠進行有意義的防禦及抗辯。

(三)系統建置者易傾向於不公開相關資料

風險評估系統的建置有其透明性的要求，不過，絕大多數的系統建置者（通常是私人企業）並不願意公開相關資訊，因為揭露系統的相關事項及蒐集公眾意見本身，會大幅地增加營運的成本¹³⁵。再者，由於商業競爭的緣故，企業也會傾向保持系統相關事項的秘密，避免其他競爭者的在技術或是所使用的資料上的仿效。最後，揭露相關資訊後，受評估者可能就可以操弄其中的事項，左右評估的結果，所以相關機關在決定要使用那一套評估系統時，可能會將資訊的公開視為是一項缺點，甚至是因而選擇其他系

¹³³ Ohio Rev. Code Ann. § 5120.114 (West Supp. 2017).

¹³⁴ See, e.g., Pa. Comm'n on Sentencing, Proposed Sentence Risk Assessment Instrument for 204 Pa. Code Chapter 305; Responses to Public Comments; Request for Proposals (Published in Pennsylvania Bulletin), available at <http://www.pacodeandbulletin.gov/Display/pabull?file=/secure/pabulletin/data/vol48/48-35/1342.html> (last visited: 2020.04.28).

¹³⁵ See Michael Mattioli, *Disclosing Big Data*, 99 MINN. L. REV. 535, 573 (2014).

元照出版提供 請勿公開散布

統¹³⁶。事實上，多數公司將建置風險評估系統的相關資訊視為營業秘密，所以會以契約保護系統的相關資訊，禁止使用的政府機關揭露¹³⁷。即使是非營利組織所開發的系統，可能也是如此¹³⁸。

風險評估系統的透明有其程序上的重要性，但基於前述種種理由，絕大多數的系統建置者並不會公開系統的相關事項，以維護自身的（商業）利益及競爭優勢，但外界也就無從得知系統是如何運作，更遑論表示意見或針對運算結果聲明不服¹³⁹。也因此，為使系統有其基本的透明性，立法者應制定必要規範，要求政府在決定使用風險評估系統時，其所使用的系統必須要公開揭露一定的事項。不過，可以預見的是，當人工智慧或是風險評估系統屬於私人企業時，其更會以人工智慧屬於封閉性軟體、保護營業秘密或維護自身市場優勢為由，拒絕揭露其相關事項¹⁴⁰。為使所用的風險評估系統遵循透明性的要求，制度上可以由立法者以民主審議的方式，彙整多方意見後，決定建置者應揭露及得不公開哪些事項或資訊¹⁴¹。此外，在立法者形成規範之前，國家機關也可以在採購時，於契約中明定必須要揭露系統中相關事項的條款。不過，可能

¹³⁶ See *id.* at 549. 在揭露相關資訊後，受評估者可能得以左右評估結果的問題，同樣地會發生在公部門所開發建置的風險評估系統中。

¹³⁷ See, e.g., *State v. Loomis*, 881 N.W.2d 749, 761 (Wis. 2016).

¹³⁸ See Nicholas Diakopoulos, *We Need to Know the Algorithms the Government Uses to Make Important Decisions About Us*, Conversation (May 24, 2016, 10:48 AM), available at <https://theconversation.com/we-need-to-know-the-algorithms-the-government-uses-to-make-important-decisions-about-us-57869> (last visited: 2020.04.27).

¹³⁹ See, e.g., *Loomis*, 881 N.W.2d at 761.

¹⁴⁰ Wexler, *supra* note 49, at 1349-50.

¹⁴¹ See, e.g., David S. Levine, *Secrecy and Unaccountability: Trade Secrets in Our Public Infrastructure*, 59 FLA. L. REV. 135, 140 (2007); Eaglin, *supra* note 64, at 114.

元照出版提供 請勿公開散布

無法避免的是，無論是以立法或是契約的方式作為透明性的手段，都可能會因為必須要公開系統的相關資訊，影響到廠商與政府合作的意願，甚至是造成無人願意投資建置類似系統，或是不願意提供此一服務給政府機關的情形。

四、透明化便足以正當化風險評估系統的使用（？）

為使外界能夠瞭解系統運作結果的原因、判斷結果是否合理，並使相對人得以聲明不服，對於風險評估系統應有其透明性的要求。事實上，透明性並不是一個全然陌生的概念，為了使某個程序能夠取信於社會大眾，國家機關常會在規範的制定過程或是在程序中，邀請外界參與，使公眾得以盡可能地知悉各個相關事項的內容。在過往，這樣的作法會有一定效果，但就風險評估系統這類型的制度，可能就不是如此。

從結論上來說，單純的「透明化」，就促進人民對於風險評估系統技術的瞭解或是信任的促進，效果可能相當有限。假設在羈押審理程序裡，法官使用了風險評估系統，確認得否釋放犯罪嫌疑人，以及決定釋放時的條件，公開了所有系統所使用的資料庫、風險因子，以及演算法等事項，被告及社會大眾很可能還是不會知道為什麼被告屬於高（或中或低）風險族群¹⁴²。其中的原因並不難理解，因為許多風險評估系統所涉及的技術或是專業，非數據專家、精熟於演算法或人工智慧的電腦工程師或是資訊學者，難以理解。再者，即使揭露前述系統的相關資訊，外界還是無從知悉許多系統中所涉及價值判斷的事項，諸如為何選擇所使用的資料庫、為什麼選擇某一些風險因子，或是為什麼風險群組間的界分點是如此等等。這些問題的答案往往比系統的建置資料本身來得重要許多，

¹⁴² Eaglin, *supra* note 64, at 106.

元照出版提供 請勿公開散布

但是往往無法從前述的資料知悉其梗概。

雖然系統相關資訊的透明不完全能夠保障當事人的程序權利，也不當然就可以確保系統運作結果的公正性及正確性，透明性仍然是使用風險評估系統時的一個重要的要求。這是因為，如果當事人及外界無從知悉系統的相關資訊，會根本沒有辦法確認系統是否使用了不當的事項或是資料，有無不合適的設定或是其中的價值判斷是否妥適。也因此，若是在刑事程序中使用了風險評估系統，還是應該有其透明性的要求，但也不能認為只要公開了系統的相關資訊，就可認為已經給予當事人完整的保障，或就足以確保系統運作結果正當。由於風險評估系統的特性，以及對於系統使用的正當性的需要，除了透明之外，系統還是必須要合於其他的要求及要件。

二、當責性

透明性是確保制度運作的一個重要的要求。在一般的情形中，若是一個制度對外不透明，外界無從知悉其重要事項，也就根本沒有辦法評估系統的使用或是結果的妥適¹⁴³。但同前所述，透明性並不足以因應實際需要，制度上還是必須要有其他的要求及要件，方能確保系統運作結果的正確性、正當性及保障受評估人的相關權益。也因此，近來有越來越多的學者主張，在刑事司法程序中使用風險評估系統之類的工具時，應要求其「當責」（accountability），而不只是單純的對外揭露相關資訊，或是透明而已¹⁴⁴。

¹⁴³ See Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1109, 1169 (2017). See also Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1533 (2013).

¹⁴⁴ See Ferguson, *id.* at 1168-71.

元照出版提供 請勿公開散布

(一)概念

一般來說，當責（性）指的是某人或某組織為其行為，向另一個人或是組織負責、解釋、說明¹⁴⁵。也有學者認為，當責性還包括了當外界對於官員或組織的行為有所評述時，能有積極的回應，如修正或調整其行為，並負起相對應的法律責任¹⁴⁶。據此，在涉及到風險評估系統的建置及使用時，當責性所指的是，評估系統的使用，必須要能夠達到起初所設定的目標，且有助於刑事司法及一般人民的利益¹⁴⁷。更為具體的作法，可以從幾個不同的面向來說明。

(二)外部及民主參與

由於風險評估系統涉及到許多價值判斷及政策的決定，並不是單純、客觀的技術或是電腦程式碼的編寫，也因此，其建置過程中應有民主參與，以確保系統能夠忠實地反映社會大眾的價值觀需要及期待。人民對於評估系統所涉及事項的價值判斷，才能夠體現於系統的運作結果¹⁴⁸。舉例來說，若是在量刑程序中容許法官參考風險評估系統的運作結果，什麼樣的風險值對應到何種刑法、刑度的高低、處遇或是釋放條件，應該由一般社會大眾、法院、法律學者或是規範制定者判斷，而不是由科學、統計、數學公式或程式設

¹⁴⁵ See Richard Mulgan, "Accountability": An Ever-Expanding Concept?, 78 PUB. ADMIN. 555, 555-56 (2000); Kimberly N. Brown, "We the People," *Constitutional Accountability, and Outsourcing Government*, 88 IND. L.J. 1347, 1350 (2013); Joanna Bird, *Regulating the Regulators: Accountability of Australian Regulators*, 35 MELB. U. L. REV. 739, 741 (2011).

¹⁴⁶ See Bird, *id.* at 741.

¹⁴⁷ Ferguson, *supra* note 143, at 1169.

¹⁴⁸ Barry Friedman & Maria Ponomarenko, *Democratic Policing*, 90 N.Y.U. L. REV. 1827, 1836 (2015).



計師決定，因為其本身是高度的價值判斷，而不是單純的數據或是公式¹⁴⁹。風險評估系統的民主或公眾參與的具體方式可以是資訊的公開及徵求建議，也可以更正式地舉行公聽會，以決定所要使用的風險因子等事項。此外，立法者也可以透過立法，規定容許評估系統使用的風險因子¹⁵⁰。

系統建置過程的外部參與，有助於使外界更加地瞭解刑事司法。由於風險評估系統的運作方式及原理，犯罪嫌疑人、被告、受判決人、律師、檢察官、法官及社會大眾，可能都無法理解為什麼特定人屬於某個再犯的風險族群¹⁵¹。風險評估系統建置過程中的公眾參與，可以讓刑事程序參與者及人民更加地瞭解到系統的相關事項及如何運作，一方面能夠使風險評估系統的使用透明化，另一方面則是促進系統使用的正當性。

(三)使用時的應告知事項

不同的風險評估系統有其不同的設置目的、功能及適用範圍的限制，並不是只要名為風險評估系統，就可以用來判斷特定人在刑事程序中的任何一種風險。舉例來說，使用再犯罪統計數據所建置的系統，很可能就不適合用來評估逃亡或不能到場的風險，也不完全適合用來判斷被告應受有的刑度高低。是故，使用風險評估系統者（法官、假釋審查委員會或是其他官員）應該要受有完整的告知，甚至是必要的培訓，以知悉系統的設置目的、程序及其限制，如此一來，才能確保其依循系統所設置的目的、範圍及功能來使用

¹⁴⁹ Christopher Slobogin, *The Civilization of the Criminal Law*, 58 VAND. L. REV. 121, 167 (2005).

¹⁵⁰ Eaglin, *supra* note 64, at 109.

¹⁵¹ See, e.g., *Loomis*, 881 N.W.2d at 774 (Abrahamson, J., concurring).

系統¹⁵²。在Loomis案中，威斯康辛州最高法院便判定，若要在量刑前調查報告中使用COMPAS之類的風險評估系統，必須要同時確保法院清楚知悉下列事項：1. COMPAS本身是封閉性程式，並未對外揭露風險因子、權重及數值如何產生；2. 風險評估及風險值是以全國為基礎進行比較及計算，而不是當地或該社區；3. 有研究顯示，COMPAS有著不當將少數族群判定為高風險的疑慮；以及4. 因為人口及族群結構不斷改變，所以風險評估程式必須要受有監督及調整，以確保其正確性¹⁵³。此外，在程序中也應告知法院，COMPAS的運算結果供矯正機關決定處遇、管束及假釋時的參考，而不是供法院作為量刑之用¹⁵⁴。

（四）獨立的查核機制及事後的評估程序

從表面上看來，風險評估系統完全是基於客觀事實及證據，但因著在建置中的種種因素，可能會存在著偏見的疑慮，例如在計算相對人再犯風險時，考量了不當的因素，如種族、國籍或宗教信仰等¹⁵⁵。也因此，原先希望可以用以降低刑度或是類似決定的演算法，可能達到完全相反的結果。例如，紐約引入了一套名為「公共安全評估」（The Public Safety Assessment）的系統，目的是為了降低不必要的審前羈押，但是由於其使用了種族等不當的變數，可

¹⁵² Ferguson, *supra* note 143, at 1167-69. See also Erik Luna, *Transparent Policing*, 85 IOWA L. REV. 1107, 1168 (2000).

¹⁵³ Loomis, 371 Wis.2d at 264.

¹⁵⁴ *Id.* at 276-77. See also Cecelia Klingele, *The Promises and Perils of Evidence-Based Corrections*, 91 NOTRE DAME L. REV. 537, 576 (2015). 在Loomis案中，威斯康辛州最高法院便判定，由於在量刑程序中，巡迴法院都已經受有這一些事項的告知，並遵循使用COMPAS時的限制，所以其所進行的量刑程序並沒有侵害被告依正當法律程序所得享有的權利。*Id.* at 277-78.

¹⁵⁵ See generally Starr, *supra* note 31, at 803.



能反而造成了反效果，增加了羈押的案件數量，使得更多人的身自由受到了不當且不必要的拘禁¹⁵⁶。

系統除了可能會有偏見或是不當的設計外，也可能存在著人所造成的（其他）錯誤。詳細地來說，在處分或是裁判中使用人工智慧或是演算法，目的之一可能是為了降低或是避免人可能發生的各種錯誤。其中的原因，並不難理解。只要是人所經手的程序，無論是什麼人，也不管是何種程序，都會存在人所造成的疏誤，如誤寫或誤算，甚至連所適用的法條，都可能會有引用上的錯誤¹⁵⁷。也因此，會有在程序中導入電腦程式或是系統輔助的倡議，幫助法官或是其他官員形成正確的決定。

電腦程式或是系統可以協助人們處理許多複雜的問題上，不過這並不是因為電腦等設備可以將複雜的事情簡單化，而是他們可以強化人們對於事物的認（感）知、處理速度及正確率，可以幫助人們看到、聽到或察覺到可能會忽略的資訊。此外，電腦也能夠在短時間內正確地處理完大量的資料。由於電腦的運算速度遠快於人，且不會有疲憊或是情緒等問題，所以在規律性、反覆性及機械性的事務中，電腦程式的表現會比人更好。然而，目前電腦程式還是必須要由人編寫，程式所需要的資料還是由人所輸入，指令還是由人

¹⁵⁶ See Roth, *supra* note 132, at 1275. Loomis案也討論到了事後監督機制。被告在該案中爭執道，一旦法院決定使用COMPAS之類的風險評估系統，就必須要有適當的監督，以確保其正確性。Loomis, 371 Wis.2d at 263-64.

¹⁵⁷ 法院判決中有誤寫或誤算等顯然錯誤者，不在少數，態樣眾多，有條文漏載者（如108年度臺上字第2531號裁定）、判決日期誤繕（108年度臺上字第1536號裁定）、辯護人姓名誤寫（107年度臺上字第3722號裁定）、上訴檢察官所屬機關誤寫（107年度臺上字第607號裁定）、不得抗告的裁定誤載為得再抗告（108年度臺抗字第286號裁定）、漏寫案件是否合於減刑規定（108年度臺抗字第94號裁定）、誤載發回法院（106年度臺上字第3110號裁定）及判決案號誤載（106年度臺職字第16號）等。

元照出版提供 請勿公開散布

下達，也因此，即使以人工智慧或是演算法來協助人來判斷，或直接用電腦作成決定，並不當然就能夠解決人會出錯的問題。亦即，人的錯誤還是會存在，只是發生在不同的地方而已。具體地來說，程式中的錯誤，很高比例都是源自於人為的編碼或是輸入¹⁵⁸。類似的錯誤，如果存在於公權力機關（如執法單位）的系統裡，通常很不容易在第一時間被發現。多半都要到數個月，甚至是數年後，有人因為系統的資訊，權利受有侵害，聲明不服，或是有軟體工程師檢查，才會察覺其中存在著人為的疏誤¹⁵⁹。在風險評估系統中，也當然地無法免除類似情況的發生。舉例來說，一套名為ASSYST，用計算依美國聯邦量刑規則所規定的刑度的程式，便曾出錯過很長一段時間，發生錯誤的主要原因是其所依據的量刑規則在一九九六年，就已修正，但程式卻沒有同步修正，以至於程式所計算出來的刑度與法規所規定的刑度有所出入¹⁶⁰。

從結論上來說，雖然系統或電腦程式可以快速地處理大量資訊，完成繁複的運算，在機械性、規律性及反覆性的事務處理上，有著遠遠超過人的表現，但系統中仍然可能有著偏見等其他不當的

¹⁵⁸ 美國聯邦最高法院就曾經審理數個涉及人為的電腦資料輸入錯誤，以及聯邦憲法第4增修條文的案件。在Florence v. Bd. of Chosen Freeholders案中，因為電腦資料的輸入錯誤，警察誤以為對Florence的逮捕令狀仍有效，於是逮捕了Florence，並將其移送至監獄。在入監前，監獄官員對Florence實施了身體檢查，其中包括脫衣、檢視體穴、傷痕、刺青及生殖器等部位（566 U.S. 318, 323-25 (2012)）。在Herring v. United States案，對於被告的逮捕令狀已經被撤銷，但是電腦資料庫並未更新，所以警察逮捕了被告，並實施附帶搜索（555 U.S. 135, 135 (2009)）。在Arizona v. Evans案中，也有著類似的事實（514 U.S. 1, 6 (1995)）。在這些案件中，警察之所以逮捕及搜索被告，都是因為電腦資料的輸入錯誤。由於警察都是善意地相信電腦資料，所以聯邦最高法院都警察所實施的強制處分合於聯邦憲法第4增修條文，並不違憲。

¹⁵⁹ Roth, *supra* note 132, at 1276.

¹⁶⁰ *Id.* at 1275. See U.S. Sent'g Comm'n, Annual Report—Year in Review (1996).



因素，或存在著其他人所造成的錯誤，如程式碼的編寫或是輸入的資料有誤。類似的錯誤若是存在於刑事程序中所使用的風險評估系統，對於受評估者的權利或人身自由就可能造成嚴重的侵害。因此，若是在刑事程序中導入風險評估系統，協助法官或其他官員作成決定，風險評估系統所使用的資料，其蒐集、分析、保存、輸入及使用，應受有獨立的查核¹⁶¹。查核機制除了應包含同儕審查（peer review）及外部檢測外，也必須要有人工智慧及資料專家學者的參與，以確保查核機制的公正性及系統運作結果與所使用的資料間的一致性¹⁶²。查核的參與人員、方法、經過及結果應保存一定期間，以供日後的查證及比對之用¹⁶³。再者，系統的建置者應一併提出可供事後監督的標準，定期檢視系統的程式碼、程式、演算法及所輸入的資料內容，確保系統是否能夠達到起初所設定的目標，並避免系統中的人為疏誤不當地侵害了相對人的人身自由等權益¹⁶⁴。監督的標準可以是客觀性、主觀性或相對性的。主觀性標準者，可以由特定人審查判斷系統是否發揮其所設定的功用；客觀性標準，如評估系統的使用可以減少受釋放者的犯罪（再犯）率。至於相對性標準，則是與官員的判斷或決定相比較，視風險評估系統對於受釋放者犯罪（再犯）的預測是否更為準確¹⁶⁵。在Loomis案中，威斯康辛州最高法院也提醒道，風險評估系統必須受有持續性監督及修正，確保其合於族群及次族群的真實景況¹⁶⁶。

¹⁶¹ Zarsky, *supra* note 143, at 1553; Ferguson, *supra* note 143, at 1167.

¹⁶² Citron, *supra* note 113, at 1305; Ferguson, *supra* note 143, at 1167.

¹⁶³ Ferguson, *supra* note 143, at 1167.

¹⁶⁴ Andrew Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion*, 62 EMORY L.J. 259, 324 (2012).

¹⁶⁵ Ferguson, *supra* note 143, at 1167-69. *See also* Luna, *supra* note 152, at 1168.

¹⁶⁶ *Loomis*, 371 Wis.2d at 276-77.

三、不應作為唯一或決定性因素

在判斷特定人是否為適於機構外處遇的低再犯風險者、對於公眾所可能造成的風險、社區處遇是否能降低再犯風險，以及評估緩刑及管束條件等情形中，風險評估系統對於協助官員形成妥適的決定，有其助益¹⁶⁷。然而，系統只是法院能夠使用的諸多工具中的一種，其結果不能是決定性的因素。其中的原因包括有，風險評估系統的建置目的或是功用，不當然包括所應用的程序的目的。例如，在Loomis案中所涉及的COMAPS，其目的或功用並未涵蓋決定被告是否應受有監禁或是刑度的長短，所納入的風險因子並不包含所有課處刑罰時所應該考慮到的事項。亦即，COMPAS僅考量了再犯，但是刑罰的目的還會有嚇阻（deterrence）、復歸（rehabilitation）、應報（retribution）及隔離（segregation）等，所以威斯康辛州最高法院提醒道，在法官量刑程序中雖然可以使用COMPAS，但其不能作為決定拘禁與否、釋放條件及刑度的唯一工具或決定性因素。這也就是為什麼威斯康辛州的矯正機關在量刑前調查報告中會特別提醒法官，風險評估系統的結果不能用來決定被告是否應受監禁、刑度長短或是社區處遇的有效性¹⁶⁸。

除了前述的理由外，另一個原因在於，被告有權受法官個別決定刑度或是依其個人情狀受有處遇。相較於前者，這一個因素更為重要。以COMPAS之類的風險評估系統為例，其係以具有與被告類似的背景或條件者的再犯可能性，計算出被告的風險值，所以系統就不是針對被告個人的情形量定其刑度及緩刑與否，而是與有近似條件的族群為對比，推測釋放後的再犯率。亦即，風險評估是就特定群體，而不是個人的行為預測。舉例來說，一個年輕、失業、

¹⁶⁷ *Id.* at 271-73.

¹⁶⁸ *Id.* at 274-75.



青少年時期曾經被捕及違反釋放條件者，即使沒有暴力犯罪紀錄，仍然可能會因為與特定族群的背景相類似，而被程式判定為具有中或高度的暴力犯罪風險¹⁶⁹。這樣一來，就可能就違背了正當法律程序，侵害了被告依其個人情狀受法院裁判決定刑度或緩刑與否的權利¹⁷⁰。是故，風險評估系統的運算結果，不能作為量刑或是關於人們人身自由限制的唯一或決定性因素，在個案中，法官及其他官員還是必須要運用自身的專業及裁量，作成妥適的判斷。在Loomis案中，威斯康辛州最高法院雖然支持下級審法院COMPAS的使用，肯認COMPAS能夠提供法院關於量刑的重要資訊，並幫助法院作成個別化及妥適的量刑，但也特別提醒道，COMPAS不能作為量刑時的唯一工具或考慮因素。為了確保法院能夠遵守此一要求，威斯康辛州最高法院特別強調道，法院對於量刑的決定負有說理的義務¹⁷¹。

四、刑罰的目的及再犯風險的評估

在使用COMPAS之類的風險評估工具時，除了要遵循前述的要件及程序外，還必須要注意到系統的建置目的及功能，以及其與刑罰的目的間是否一致。詳細地來說，在討論到刑罰的目的或是基礎理論時，多會論及應報（retribution）；威嚇或嚇阻（deterrence）；再教育或社會復歸（rehabilitation），以及隔離（incapacitation）¹⁷²。立法者通常不會將刑罰限制在單一目的，而是會併合數個不同的理論基礎，用以說明或解釋課予受判決人刑罰的原因，並使外界得以

¹⁶⁹ *Id.* at 265-66.

¹⁷⁰ *Id.* at 265.

¹⁷¹ *Id.* at 275.

¹⁷² See, e.g., Aya Gruber, *A Distributive Theory of Criminal Law*, 52 WM. & MARY L. REV. 1, 4 (2010); Casey et al., *supra* note 30, at 1.

元照出版提供 請勿公開散布

檢視國家所課予被告的刑罰是否妥適¹⁷³。

刑罰目的的設定會影響到風險評估系統的使用。亦即，要在量刑程序中使用風險評估系統時，系統的建置目的或是功能就必須要與立法者所設定的刑罰目的一致，否則系統的運算結果對於決定刑度的參考價值，就相當有限。舉例來說，若風險評估系統是用來確認特定人逃亡及是否能到場的可能性，而刑罰的主要目的設定在應報或是社會復歸時，使用系統的運作結果來決定刑罰，可能就會很大的出入。相對地，若認為刑罰的目的或原因在於，特定人對他人或社會帶有危險性，為了避免其造成危害，所以將其隔絕於社會共同之外¹⁷⁴，而風險評估系統是用以判斷受測者將來的犯罪可能性時，在量刑程序中，就可以使用該評估系統來決定其所應受有的刑罰、處遇或釋放條件，以避免或降低被告對他人或公眾所可能造成的危害¹⁷⁵。

¹⁷³ Michael Tonry, *Purposes and Functions of Sentencing*, 34 CRIME & JUST. 1, 10 (2006).

¹⁷⁴ ARTHUR W. CAMPBELL, LAW OF SENTENCING § 2:3 (2019).

¹⁷⁵ See Gruber, *supra* note 172, at 4. See also Pamela M. Casey et al., Nat'l Ctr. for State Courts, Using Offender Risk and Needs Assessment Information at Sentencing: Guidance for Courts from a National Working Group 8-10 (2011), available at <http://www.ncsc.org/~media/Microsites/Files/CSI/RNA%20Guide%20Final.ashx> (last visited: 2019.12.16). 風險評估系統是否可以作為量刑的參考之用，取決於評估系統所採用的風險因子。這是因為，依隔離理論，法院所要判斷的是什麼樣的刑罰、處遇或釋放條件，可以有效地降低或是減少被告將來犯罪的可能性（See Starr, *supra* note 31, at 855-58.）。亦即，被告所受的刑罰種類及刑度本身就是一個變數，而不是系統運作的結果。但是，目前主要的風險評估系統並沒有將量刑作為量度風險，或是被告再犯可能性的變數之一（See Eaglin, *supra* note 64, at 100.）。也因此，即使是判斷再犯可能性的風險評估系統，其運算結果是否能作為刑度或處遇決定的參考，便有待商榷。

五、自動化依賴的預防

不能否認的是，在部分案件中，電腦程式、技術或是各樣科技（如DNA）的使用，可以幫助法院正確地認定事實，但有些時候，其也可能造成法官過度依賴科技，忽略了自身仍應判斷證據的證明力或技術或設備可能存有錯誤，而逕自以電腦程式或科技設備運作的結果來認定事實。當法院所使用的技術或是設備有錯誤時，其裁判也就因而所有差錯。這種使用科技所產生的疏誤，多稱之為「自動化依賴」（automation complacency）。

「自動化依賴」指的是，當人們發現自動化系統總是正確，覺得其結果是可信賴的時候，就可能會過度信任系統，認定其判斷都是正確的，因而忽略了系統結果可能存有錯誤，必須要由人進行交叉比對或驗證¹⁷⁶。關於自動化依賴，最典型的例子之一，是非常普及的導航。在沒有導航系統之前，人們會（必須）認識到各種可能的路線，不同時間的車流量、路況及道路寬窄等事項，並據以決定行車的路線，但是一旦開始使用導航系統後，即使大多數的製造商都會提醒駕駛人還是要注意道路情狀，根據當時情形作成合適的判斷，但大部分的駕駛人還是會依照系統的指示，不再臨機應變或主動地參與在路線的決定中¹⁷⁷。導航系統造成塞車或是事故的情形也就時有所聞¹⁷⁸。這種對於系統的過度依賴，不只會發生在日

¹⁷⁶ See Raja Parasuraman & Dietrich H. Manzey, *Complacency and Bias in Human Use of Automation: An Attentional Integration*, 52 HUMAN FACTORS 381, 381 (2010).

¹⁷⁷ Roth, *supra* note 132, at 1279.

¹⁷⁸ E.g., Michelle Lou, *Near 100 Drivers Followed a Google Maps Detour—and Ended Up Stuck in an Empty Field*, CNN (June 27, 2019), available at <https://edition.cnn.com/2019/06/26/us/google-maps-detour-colorado-trnd/index.html> (last visited: 2019.11.20). 另外可以參照蔡政珉，導航變搗亂！載女友去泡湯 險衝山谷前輪懸空，自由時報（2019年1月2日），<https://news.ltn.com.tw/news/>

元照出版提供 請勿公開散布

常生活中，也常見於專業領域裡，如指紋辨識。進行指紋辨識時，是由電腦系統先搜尋出數個最有可能一致的指紋，再由鑑識人員比對後，作成判讀。亦即，系統只是提供最接近的樣本，鑑識人員還是必須要自行判斷一致與否。但是，類似系統的導入，會使得鑑識人員認為（假定）系統運作的結果中，一定有一個與採樣的指紋相符，但事實並非如此。辨識系統只是給出資料庫中近似度最高的前幾個結果，但真正一致的，有可能是近似度略低的樣本，所以最後還是需要鑑識人員以其知識及經驗，做出妥適的判斷¹⁷⁹。一旦鑑識人員過度依賴辨識系統，鑑定的結果就會有其誤差，連帶地影響到法院判決的正確性。對於科技設備或是儀器的依賴，可以再以醉酒的判斷為例來說明。當提出於法院的是警察所實施的酒醉測試，如平衡動作或是倒背英文字母表時，法官通常比較會（能夠）從自身的經驗注意到測試的過程及結果中，可能存在的違失或偏見。但相對地，當出現在法庭的不是警察官員的證詞或是其所實施的測驗，而是呼氣酒精測試儀的檢測結果時，審判者多難以確認其是否正確，傾向於相信科技的客觀、公正及正確，所以會接受其作為證據之用。但也因為如此，若是機器或設備的測試結果有誤，通常都要累積有相當數量的案件，且有當事人聲明不服後，才會發現。舉例來說，在俄亥俄州，一群被判定酒後駕車的被告發現，某款呼氣酒精測試儀判定呼氣酒精濃度過高的比例過高，繼續調查後才知道，該測試儀無法正確的判斷所偵測到的酒精是來自於口腔，還是從肺部而來，所以檢測的結果有著大量的偽陽性反應¹⁸⁰。之後，俄亥俄州的法院判定該檢測儀的結果沒有證據能力，部分州也才開

Miaoli/breakingnews/2659770，最後瀏覽日：2019年11月20日。

¹⁷⁹ See Simon A. Cole, *More Than Zero: Accounting for Error in Latent Fingerprint Identification*, 95 J. CRIM. L. & CRIMINOLOGY 985, 1064-65 n.394 (2005).

¹⁸⁰ See Roth, *supra* note 132, at 1271-72.

始禁止在執法中使用該測試儀¹⁸¹。

是故，在刑事程序中導入風險評估系統後，很可能也會發生前述的自動化依賴。詳細地來說，一旦在原本是由人處理或是決定的程序中，導入電腦或設備的協助，人們就會開始傾向依賴程式，並減少人自己的判斷，因而產生自動化依賴的現象。過度地信任系統的結果，就是無法主動且即時地發現及修正其中可能的錯誤。同樣地，一旦開始使用風險評估系統，即使是作為輔助或參考之用，仍然可能使得法官或是官員傾向於依賴系統的運算結果，不再依其專業或是職權判斷相對人的再犯可能性、隱匿或逃亡的風險¹⁸²。形式上來看，決定的過程借助了電腦系統，更為客觀公正，但實質上卻是應作成實質決定或判斷的官員，隱身在數據之後，委棄了自身所應付有的職責¹⁸³。

為了防止或降低自動化依賴的情形，在刑事程序中導入風險評估系統等電腦程式時，應該要將自動化依賴的現象納入考量¹⁸⁴。舉例來說，在量刑程序中使用拘禁與否或刑度的建議系統時，應該由法官或其他官員先以傳統的方式作成初步的判斷，再由系統產生相關建議及數值，以確保官員會運用其判斷、獨立地審查及決定被告是否應受監禁、應否釋放、釋放的條件及刑度等等事項，而不是

¹⁸¹ See Alan Johnson, Judge Finds Breathalyzer Not Scientifically Reliable, Columbus Dispatch (Aug. 22, 2013, 9:08 AM), available at <http://www.dispatch.com/content/stories/local/2013/08/22/judge-finds-breathalyzer-not-scientifically-reliable.html> (last visited: 2019.09.11).

¹⁸² Starr, *supra* note 31, at 866-68.

¹⁸³ See Máximo Langer, *Rethinking Plea Bargaining: The Practice and Reform of Prosecutorial Adjudication in American Criminal Procedure*, 33 AM. J. CRIM. L. 223, 279 (2006).

¹⁸⁴ NICHOLAS CARR, *THE GLASS CAGE: AUTOMATION AND US* 164-65 (2014).

完全地盲從系統運作的結果¹⁸⁵。如此一來，才不會使得法官等官員變成電腦系統運算結果的橡皮圖章，也才能確保實際作成決定的仍然是依法負有職責的法官等官員，而不是電腦程式或系統¹⁸⁶。

陸、結 論

由於電腦、演算法及資料科學的飛快進步，人工智慧趨於成熟，不再是只能在實驗室中紙上談兵的理論，而是已經具應用性的技術。此一技術除了在人們的一般生活中開始展露頭角外，部分國家已經將之應用於刑事程序中，協助（或甚至是替代）法官或其他官員作成各種決定，諸如以人工智慧建置風險評估系統，預測特定人再犯或是逃亡的可能性，並進一步決定拘禁、釋放或釋放條件。不過，在刑事程序中應用前所未見的風險評估系統，也引起了諸多的討論及正反意見。

深入探究後可以知道，風險評估系統的建置，從資料的蒐集、資料庫的選擇、再犯概念的界定、影響再犯的因素、是否及如何修改資料中的錯誤、選擇哪些風險因子來建構預測模型、所納入的風險因子是否及如何加權、如何取得及輸入個案數據、風險群組的建立及正確性的評估等事項，都取決於人的價值判斷及主觀決定，並非

¹⁸⁵ See Jennifer Skeem, *Risk Technology in Sentencing: Testing the Promises and Perils*, 30 JUSTICE Q. 297, 299 (2013); Katherine Y. Barnes, *Against Judgment*, 93 CORNELL L. REV. 689, 697 (2008).

¹⁸⁶ 綜前所述，國家機關在刑事程序中所使用的風險評估系統，如果不具備透明性、當責性及合目的性等，應認該評估結果不合於對於科學證據的要求，不能作為決定的依據。另外，由於評估系統所產生的評估結果，是電腦就所接收到資料，依已經編寫好的程式，計算出風險數值，這個過程，就像是測速照相機偵測到超速的車輛，拍攝下照片一樣，都不是人的供述，所以不會有傳聞法則的適用，也就不會有風險評估結果屬於那一種傳聞法則例外的問題。



元照出版提供 請勿公開散布

如想像般，是純然由電腦所計算及產出的客觀、中立及全無偏見。此外，由於人工智慧的技術所致，外界幾乎無法得知風險評估系統產出其結果的原因，然而為了確認系統中可能存在的不當事項或歧視（如種族或宗教信仰）、運算結果是否合理、知悉人們可能沒有注意到的規律或現象，以及使得人們能夠提起有實質意義的救濟，採用風險評估系統的條件及程序，也應該有其透明性的要求。亦即，制度上必須要揭露系統的重要資訊及採用系統的標準與程序等事項。然而，值得注意的是，為了維護自身的商業利益的緣故，建置者（特別是私人企業）會傾向於保持不公開的狀態。再者，透明性固然是使用風險評估系統的重要基礎，但由於人工智慧技術的特性的緣故，僅僅公開一定事項，並不足以正當化風險評估系統的使用。除了透明性外，法制上還應該要有當責性的要求，確保系統能夠達到所設定的目標，運作結果的正確性及保障受評估人的相關權益。具體地來說，系統的建置過程應有外部及民主參與、應充分告知使用系統的法官及官員與系統相關的重要事項，且應建立獨立的查核機制及事後評估程序。

由於系統建置的目的，以及被告受法官個別決定刑度或個別處遇的權利等因素，風險評估系統的運算結果不應作為決定的唯一或決定性因素。於刑事程序中導入風險評估系統時，必須要注意到系統的建置目的及功能及程序的目的間是否一致，此外，在使用系統時，還應該要有相對應的機制，避免自動化依賴的情形產生。

在刑事程序中使用人工智慧所建置的風險評估系統，看似只會在科幻小說或電影中才會出現的情節，但其實已經發生在許多國家，類似的技術應用在我國的司法實務上的速度，會遠遠超過我們所能想像。這一篇稿件針對這一個議題提出些許思索後的想法，冀希能夠為相關研究，累積研究的成果及素材，也就將來實務上可能會有問題，提供些許討論的方向。



元照出版提供 請勿公開散布

參考文獻

一、中 文

1. 內政部警政署，107年警政統計年報，2019年7月。
2. 王兆鵬，重新思維測謊之證據能力，月旦法學雜誌，135期，頁137-152，2006年8月。
3. 王兆鵬、張明偉、李榮耕，刑事訴訟法（上），5版，2020年3月。
4. 司法院，刑事案件量刑審酌事項參考手冊，2018年8月。
5. 吳巡龍，科學證據與測謊的證據能力，月旦法學教室，38期，頁95-105，2005年12月。
6. 李榮耕，初探刑事程序法的人工智慧應用，載：人工智慧相關法律議題芻議，頁117-148，2018年11月。
7. 邱文聰，初探人工智慧中的個資保護發展趨勢與潛在的反歧視難題，載：人工智慧相關法律議題芻議，頁149-175，2018年11月。
8. 陳玉書，再犯特性與風險因子之研究：以成年假釋人為例，載：刑事政策與犯罪研究論文集，16期，頁1-26，2013年12月。
9. 曾春僑，儀器測謊與常見疑義，在野法潮，46期，頁74-79，2020年7月。
10. 黃虹霞，測謊報告與證據能力、證明力——一個執業律師的看法，萬國法律，149期，頁31-36，2006年10月。
11. 劉靜怡，人工智慧潛在倫理與法律議題鳥瞰與初步分析，載：人工智慧相關法律議題芻議，頁1-45，2018年11月。

二、外 文

1. Alder, Ken (2007), THE LIE DETECTORS: THE HISTORY OF AN AMERICAN OBSESSION, NY: Simon & Schuster.
2. Barnes, Katherine Y., *Against Judgment*, 93 CORNELL L. REV. 689 (2008).
3. Barocas, Solon & Selbst, Andrew D., *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016).

4. Bird, Joanna, *Regulating the Regulators: Accountability of Australian Regulators*, 35 MELB. U. L. REV. 739 (2011).
5. Brown, Kimberly N., "We the People," *Constitutional Accountability, and Outsourcing Government*, 88 IND. L.J. 1347 (2013).
6. Calo, Ryan, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399 (2017).
7. Campbell, Arthur W. (2019), *LAW OF SENTENCING*, Tor: Thomson/Reuters.
8. Carr, Nicholas (2014), *THE GLASS CAGE: AUTOMATION AND US*, NY: W. W. Norton & Company .
9. Cheng, Edward K. & Alexander G., *Beyond the Witness Bringing a Process Perspective to Modern Evidence Law*, 97 TEX. L. REV. 1077 (2019).
10. Citron, Danielle Keats, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2008).
11. Citron, Danielle Keats & Pasquale, Frank, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014).
12. Coglianese, Cary & Lehr, David, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1 (2019).
13. Cole, Simon A., *More Than Zero: Accounting for Error in Latent Fingerprint Identification*, 95 J. CRIM. L. & CRIMINOLOGY 985 (2005).
14. Deeks, Ashley, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLMN. L. REV. 1829 (2019).
15. Dorlaque, Joseph, *New Machines Get Drunks off the Road*, POPULAR SCI. MONTHLY 165 (1955).
16. Dror, Itiel, *The Ambition to Be Scientific: Human Expert Performance and Objectivity*, 53 SCI. & JUSTICE 81 (2013).
17. Eaglin, Jessica M., *The Drug Court Paradigm*, 53 AM. CRIM. L. REV. 595 (2016).
18. Eaglin, Jessica M., *Constructing Recidivism Risk*, 67 EMORY L.J. 59 (2017).
19. Ferguson, Andrew Guthrie, *Predictive Policing and Reasonable Suspicion*, 62 EMORY L.J. 259 (2012).
20. Ferguson, Andrew Guthrie, *Big Data and Predictive Reasonable Suspicion*, 163 U.

元照出版提供 請勿公開散布

- PA. L. REV. 327 (2015).
21. Ferguson, Andrew Guthrie (2017), *THE RISE OF BIG DATA POLICING*, NY: NYU Press.
22. Ferguson, Andrew Guthrie, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1109 (2017).
23. Friedman, Barry & Ponomarenko, Maria, *Democratic Policing*, 90 N.Y.U. L. REV. 1827 (2015).
24. Gouldin, Lauryn P., *Defining Flight Risk*, 85 U. CHI. L. REV. 677 (2018).
25. Gruber, Aya, *A Distributive Theory of Criminal Law*, 52 WM. & MARY L. REV. 1 (2010).
26. Hamilton, Melissa, *Adventures in Risk: Predicting Violent and Sexual Recidivism in Sentencing Law*, 47 ARIZ. ST. L.J. 1 (2015).
27. Hamilton, Melissa, *Back to the Future: The Influence of Criminal History on Risk Assessments*, 20 BERKELEY J. CRIM. L. 75 (2015).
28. Harcourt, Bernard E. (2007), *AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE*, Chicago: U. Chicago Press.
29. Harcourt, Bernard E., *Risk as a Proxy for Race: The Dangers of Risk Assessment*, 27 FED. SENT'G REP. 237 (2015).
30. Hyatt, Jordan et al., *Reform in Motion: The Promise and Perils of Incorporating Risk Assessments and Cost-Benefit Analysis into Pennsylvania Sentencing*, 49 DUQ. L. REV. 707 (2011).
31. Johnson, James L. et al., *The Construction and Validation of the Federal Post Conviction Risk Assessment*, 75-SEP FED. PROBATION 16 (2011).
32. Kern, Richard P. & Farrar-Owens, Meredith, *Sentencing Guidelines with Integrated Offender Risk Assessment*, 16 FED. SENT'G REP. 165 (2004).
33. Klingele, Cecelia, *The Promises and Perils of Evidence—Based Corrections*, 91 NOTRE DAME L. REV. 537 (2015).
34. Koepke, John Logan & Robinson, David G., *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725 (2018).
35. Kroll, Joshua A. et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633 (2017).

36. Langer, Máximo, *Rethinking Plea Bargaining: The Practice and Reform of Prosecutorial Adjudication in American Criminal Procedure*, 33 AM. J. CRIM. L. 223 (2006).
37. Levine, David S., *Secrecy and Unaccountability: Trade Secrets in Our Public Infrastructure*, 59 FLA. L. REV. 135 (2007).
38. Luna, Erik, *Transparent Policing*, 85 IOWA L. REV. 1107 (2000).
39. Maloney, Douglas C., *Lies, Damn Lies, and Polygraphs: The Problematic Role of Polygraphs in Postconviction Sex Offender Treatment (PCSOT)*, 84 TEMP. L. REV. 903 (2012).
40. Mattioli, Michael, *Disclosing Big Data*, 99 MINN. L. REV. 535 (2014).
41. Mayson, Sandra G., *Dangerous Defendants*, 127 YALE L.J. 490 (2018).
42. McCarthy, John et al., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence (Aug. 31, 1955)*, 27(4) AI MAGAZINE 12 (2006).
43. Mindell, David A. (2015), *OUR ROBOTS, OURSELVES: ROBOTICS AND THE MYTHS OF AUTONOMY*, NY: Viking.
44. Monahan, John, *A Jurisprudence of Risk Assessment: Forecasting Harm Among Prisoners, Predators, and Patients*, 92 VA. L. REV. 391 (2006).
45. Monahan, John & Skeem, Jennifer L., *Risk Assessment in Criminal Sentencing*, 12 ANN. REV. CLINICAL PSYCHOL. 489 (2016).
46. Mulgan, Richard, *"Accountability": An Ever-Expanding Concept?*, 78 PUB. ADMIN. 555 (2000).
47. Murphy, Erin, *The Mismatch Between Twenty-First-Century Forensic Evidence and Our Antiquated Criminal Justice System*, 87 S. CAL. L. REV. 633 (2014).
48. Oleson, J. C., *Risk in Sentencing: Constitutionally Suspect Variables and Evidence-Based Sentencing*, 64 SMU L. REV. 1329 (2011).
49. Parasuraman, Raja & Manzey, Dietrich H., *Complacency and Bias in Human Use of Automation: An Attentional Integration*, 52 HUMAN FACTORS 381 (2010).
50. Quinsey, Vernon L. et al. (2006), *Violence Risk Appraisal Guide (VRAG)*, in Vernon L. Quinsey, Grant T. Harris, Marnie E. Rice & Catherine A. Cormier eds., *VIOLENT OFFENDERS: APPRAISING AND MANAGING* (2d ed.). (D.C.: Amer. Psychological Assn.).

- 元照出版提供 請勿公開散布
51. Remus, Dana A., *The Uncertain Promise of Predictive Coding*, 99 IOWA L. REV. 1691 (2014).
52. Roth, Andrea, *The Uneasy Case for Marijuana as Chemical Impairment Under a Science-Based Jurisprudence of Dangerousness*, 103 CAL. L. REV. 841 (2015).
53. Roth, Andrea, *Trial by Machine*, 104 GEO. L.J. 1245 (2016).
54. Scurich, Nicholas, *Evidence-Based Sentencing: Public Openness and Opposition to Using Gender, Age, and Race as Risk Factors for Recidivism*, 40 LAW & HUM. BEHAV. 36 (2016).
55. Sidhu, Dawinder S., *Moneyball Sentencing*, 56 B.C. L. REV. 671 (2015).
56. Sites, Brian, *Rise of the Machines: Machine-Generated Data and the Confrontation Clause*, 16 COLUM. SCI. & TECH. L. REV. 36 (2014).
57. Skeem, Jennifer, *Risk Technology in Sentencing: Testing the Promises and Perils*, 30 JUSTICE Q. 297 (2013).
58. Sklansky, David Alan, *Hearsay's Last Hurrah*, 2009 SUP. CT. REV. 1 (2009).
59. Slobogin, Christopher, *The Civilization of the Criminal Law*, 58 VAND. L. REV. 121 (2005).
60. Slobogin, Christopher, *Dangerousness and Expertise Redux*, 56 EMORY L.J. 275 (2006).
61. Starr, Sonja B., *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803 (2014).
62. Tene, Omer & Polonetsky, Jules, *Judged by the Tin Man: Individual Rights in the Age of Big Data*, 11 J. TELECOMM. & HIGH TECH. L. 351 (2013).
63. Tonry, Michael, *Purposes and Functions of Sentencing*, 34 CRIME & JUST. 1 (2006).
64. Tonry, Michael, *Legal and Ethical Issues in the Prediction of Recidivism*, 26 FED. SENT'G REP. 167 (2014).
65. Weisberg, Robert, *Meanings and Measures of Recidivism*, 87 S. CAL. L. REV. 785 (2014).
66. Wexler, Rebecca, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018).
67. Zarsky, Tal Z., *Transparent Predictions*, 2013 U. ILL. L. REV. 1503 (2013).



元照出版提供 請勿公開散布

AI Risk Assessment in Criminal Justice

Rong-Geng Li^{*}

Abstract

Due to the increasing maturity of artificial intelligence (AI), several countries (e.g., the United States) have applied such technology to construct risk assessment system as a way to predict certain people's risk of committing crimes and decide whether they should be detained or released. After thorough analysis, however, it is clear that the construction of risk assessment is not as just or objective as people might think. People's subjective decisions and value judgment exist in every stage of the construction of system. In addition, due to the nature of AI, the decision-making processes of the risk assessment system is often unclear, meaning that the system lacks transparency. Therefore, we suggest that such risk assessment system must comply with the transparency requirement, and the use of system should be in line with the accountability requirement to ensure the system meets the set goal, the accuracy of the of its outcome, and people's rights. Furthermore, the outcome of the risk assessment system should not be the determinative factor of the decisions in criminal procedure, and whether the function of the system is consistent with the purpose of the process should be

^{*} Professor of Law, National Taipei University Law School; S.J.D., Indiana University—Bloomington Maurer School of Law.

Received: January 22, 2021; accepted: July 25, 2021

 元照出版提供 請勿公開散布。
verified. Finally, there should be effective mechanism to avoid automation complacency.

Keywords: Artificial Intelligence, Risk, Risk Assessment System, Transparency, Accountability, Automation Complacency