

論刑事算法開示制度

張培錚*

摘要 刑事訴訟中算法存在遮蔽式運用，表現為偵查過程中算法的雙重黑箱、鑒定制度中算法認知隱而不彰、庭審過程中算法證據的概括式審查。算法遮蔽式運用，可能會誘發算法內容封閉化與零散化、技術性正當程序缺失、助長司法工作人員的依賴性等風險。有必要構建形式算法開示制度，促進統一完善的標準形成、奠定刑事技術性正當程序基礎以及促進算法問責。算法開示的構建需要遵循兩種邏輯：依托職權、服務於訴訟效率的權力行使邏輯和保障訴訟參與、接受公正審判權的權利保障邏輯。算法開示構建過程中需要妥善處理好算法開示與商業、國家秘密的沖突、算法複雜性導致無效開示、傳統開示制度都對算法開示的阻礙等問題。構建刑事算法開示制度需要明確權利保障優先原則、析理開示為主原則、強制開示原則以及全程開示原則，從算法開發環節、立案、偵查、審查起訴、審判等階段落實相關原則，構建具體程序。

關鍵詞 算法開示 算法公開 析理開示 技術性正當程序

一、問題的提出

算法廣泛而深刻的介入刑事司法，改變了刑事司法實踐樣態。其中，算法、大數據技術等新興科技嵌入到偵查、鑒定等訴訟活動之中，幫助司法工作人員收集、組織、分析信息，並形成相應的證據。偵查過程中，司法工作人員利用算法識別犯罪嫌疑人。例如，美國國會暴動案後，FBI 利用面部識別程序，鎖定犯罪嫌疑人身份。^[1] 鑒定過程中，技術人員利用算法賦能鑒定技術，高效得出分析報告。例如，採用 FST、TrueAllele 等算法分析程序，分析複雜 DNA 混合物。

盡管，算法扮演了至關重要的角色，但算法通常並不向辯方公開。遮蔽式運用對傳統刑事訴訟模式形成了巨大沖擊，克減了犯罪嫌疑人、被告人的訴訟權利。鑒於算法在生成證據過程中發揮著核心作用，有必要設計專門的訴訟程序對算法進行公開，筆者將這一訴訟程序稱之為“算法開

* 張培錚，中國政法大學司法文明協同创新中心博士研究生。

本文系教育部人文社會科學研究規劃項目“區塊鏈證據規則研究”（項目編號：22YJA820034）階段性成果。

[1] Ryan Reilly, *Feds Track Down Capitol Rioter With Facial Recognition Hit on His Girlfriend's Instagram*, HuffPost (20/04/2021), https://www.huffingtonpost.co.uk/entry/facial-recognition-capitol-defendants_n_607f34c0e4b0df3610c17614?ri18n=true.

示”。但我國目前尚未存在發揮算法進行開示功能的制度，具有類似功能的證據開示、辯護制度在數字時代下也難以平衡或補強犯罪嫌疑人、被告人的弱勢地位，^[2]基於此，有必要探索構建我國刑事算法開示制度。

當前，學界對算法開示制度欠缺專門性的探討，有關討論散見於其他主題研究之中。總體而言，既有研究對算法開示相關問題的討論主要集中於：第一，算法開示制度作為大數據偵查、證明、證據審查等保障措施。學者認為，程序透明也是大數據證明可靠與否的重要因素，算法開示等透明機制有利於提升數據質量、算法準確性。^[3]也有學者認為，通過算法開示制度規制大數據偵查，降低其風險。^[4]還有學者認為，審查大數據證據的可靠性，需要設置算法科學性規則，其中涉及算法開示規則，並指出算法開示需要符合必要性、針對性、嚴格保密措施以及強制性排除規則等要求。^[5]也有學者對此提出相反意見，他認為保障人工智能證據的審查，並不需要開示算法。^[6]第二，算法開示是刑事證據開示在數字時代下的新發展。有學者認為，刑事開示制度面臨算法不公開的挑戰，使得運作效果大打折扣，有必要對算法進行適度公開。^[7]第三，探討算法開示制度構建可能遇到的障礙。有學者認為，算法開示制度當前面臨障礙是控方為追訴利益不願公開算法以及科技企業從商業利益和知識產權保護角度拒絕算法公開。^[8]也有學者認為，構建算法開示制度需要聚焦算法透明與商業秘密保護的沖突問題。^[9]第四，構建算法開示的具體制度。有學者建議從開示環節、開示保障以及專家出庭制度等三個方面設置具體開示程序；^[10]也有學者認為算法開示制度不應有法律進行細致規定，應當由控辯雙方進行協商，若出現爭議，則交由法官確定。^[11]

上述理論爭鳴中，可以觀察到當前學界對算法開示的探討尚有如下不足：首先，欠缺對算法開示制度的專門性研究。學界研究大都將算法開示制度視作其他制度研究的保障措施，欠缺獨立地研究算法開示制度的價值。其次，對我國刑事訴訟中算法運行現狀欠缺必要的總結。這也導致了必要性和正當性討論的欠缺。再次，算法開示制度可能遇到障礙的討論缺乏系統性。最後，現有研究出發點大都基於算法開示保障權利的應然維度，忽視了實踐中權力行使之維度，而現實制度通常需要二維共向發力，共同組成制度構建邏輯。

基於此，本文將概括實踐中算法的遮蔽式運用的樣態，指明算法遮蔽式運用風險，闡明構建算法開示制度的必要性，結合實然維度與應然維度，從法理層面分析算法開示的構建邏輯。同時，指明未來構建算法開示制度可能遇到的障礙，並基於構建原則，構建具體的算法開示制度。

[2] 參見裴煒：《論刑事數字辯護：以有效辯護為視角》，載《法治現代化研究》2020年第4期，第126頁；鄭曦：《數字時代刑事證據開示制度之重塑》，載《華東政法大學學報》2023年第4期，第41頁。

[3] 參見王燃：《大數據證明的機理及可靠性探究》，載《法學家》2022年第3期，第65頁。

[4] 魏俊斌：《大數據偵查的風險防控與“意識”匡正》，載《大連理工大學學報（社會科學版）》2024年第4期，第95頁。

[5] 洪濤：《大數據證據真實性審查規則的建構》，載《蘇州大學學報（法學版）》2024年第1期，第78頁。

[6] 楊菲：《風險與應對：論人工智能證據在刑事司法中的運用——以人臉識別分析結果為例》，載《山東警察學院學報》2022年第2期，第54頁。

[7] 鄭曦：《數字時代刑事證據開示制度之重塑》，載《華東政法大學學報》2023年第4期，第46頁。

[8] 同上註。

[9] 王燃、孫藝桐：《人身危險性評估的算法治理——從算法透明與商業秘密沖突展開》，載《上海政法學院學報》2023年第3期，第95頁。

[10] 王燃：《大數據證明的機理及可靠性探究》，載《法學家》2022年第3期，第70頁。

[11] 鄭曦：《數字時代刑事證據開示制度之重塑》，載《華東政法大學學報》2023年第4期，第47頁。

二、刑事訴訟中算法遮蔽式運用的實踐圖景

刑事訴訟實踐中，算法存在遮蔽式運用的現象，具體包括：偵查過程中算法的雙重黑箱、鑒定制度中算法認知隱而不彰、庭審過程中算法證據的概括式審查。算法遮蔽式運用是對算法在刑事司法實踐的理論概括，然而，它本身存在一定的風險，會帶來一系列的理論和實踐問題。

（一）算法遮蔽式運用的表現形式

1. 偵查過程中算法的雙重黑箱

偵查過程本身具備封閉性，算法也表現出黑箱性，兩者疊加，呈現出技術封閉與法律封閉共存的“雙重黑箱”樣態。也有學者將其稱之為“雙重封閉”，但僅道明其危害，卻沒有指出具體運作樣態。^[12]筆者對此概括如下：

算法開發封閉性。算法開發的現狀可以從微觀與宏觀兩種角度進行闡釋。從微觀上，有關算法開發呈現出“偵查—技術提供方”弱二元封閉模式。所謂“弱二元封閉模式”是指偵查機關通過經驗提出建設要求以及提供相應數據，技術提供方負責設計相應算法。^[13]微觀上二元主體之間並非完全獨立。作為使用主體、出資方的偵查機關對技術提供方具有支配地位。設計出算法並非完全價值中立，其至少包含警察部門與設計人員的主觀判斷。從宏觀上，算法開發呈現出“算法孤島”模式。所謂“算法孤島”模式是指各地公安機關和不同警種間，自主招標研發大數據偵查工具，有利於各地區各警種實際需要，設計相應的算法框架並建立相應的數據庫。^[14]設計出算法固然照顧到各個地方實際情況以及各警種實際業務開展，但散亂的開發與管理導致開發標準不統一。

算法取證的封閉性。首先，算法取證的行為性質不明，法律規制缺位，導致算法秘密運行。^[15]偵查過程中，由於算法取證行為具有“累加性基本權利幹預”^[16]的特征，學界通說將其歸為強制性偵查措施。^[17]當前我國法律並無針對算法取證的明文規定，實踐中算法取證持續秘密運行。其次，我國傳統偵查措施司法審查缺位，控制手段多為偵查機關內部自控，欠缺必要的外部制約。最後，算法取證不留痕跡。算法取證所得證據材料往往用於獲取案件線索以及作為補強證據使用，偵查機關並不會將利用算法取證過程寫入案卷之中。無論是後續審查起訴以及審判階段亦或是律師閱卷，都不會覺察發現相應的取證違法行為。

2. 鑒定制度中算法認知隱而不彰

鑒定制度具有將專業人員的認知開示的功能。不同於普通問題，專門性問題往往難以被常人

[12] 參見李金峰、陳如超：《大數據警務風險及其規制》，載《中國人民公安大學學報（社會科學版）》2023年第5期，第135頁。

[13] 參見沈國琴、齊小力：《人工智能嵌入預測警務的法律風險及其預防》，載《廣西社會科學》2021年第5期，第14頁。

[14] 參見漆晨航、陳剛：《大數據偵查算法的規制：以檢察監督為思路》，載《中國人民公安大學學報（社會科學版）》2022年第6期，第69頁。

[15] 筆者此處所言算法取證與大數據偵查內涵一直。大數據偵查是指通過計算機技術對存儲於網絡與計算機系統的海量數據進行收集、共享、清洗、比對和挖掘，從而發現犯罪線索、證據信息和犯罪嫌疑人的偵查措施與方法。筆者使用算法取證旨在強調算法（計算機技術）在此過程中的核心作用。

[16] 褚僑、施鵬鵬：《德國刑事訴訟中大數據證據合法性問題之探討》，載《人民檢察》2024年第7期，第69頁。

[17] 參見張可：《大數據偵查措施程控體系建構：前提、核心與保障》，載《東方法學》2019年第6期，第89頁；程雷：《大數據偵查的法律控制》，載《中國社會科學》2018年第11期，第157頁；胡銘、龔中航：《大數據偵查的基本定位與法律規制》，載《浙江社會科學》2019年第12期，第15頁。

所理解。鑒定制度作為解決專門性問題重要途徑，具備將專業問題或技術人員的專業知識開示的功能。例如，美國梅倫德斯-迪亞茲案將化學技術人員分析書歸類於“證言性”陳述，受到憲法第六修正案制約，需要給予被告人與其對抗的機會，這使得專門性知識在對抗制法庭中得以開示。又如，德國《刑事訴訟法》第85條規定了鑒定專家適用證人條款；而第48條第1款，規定了證人到場以及陳述的義務，這也使得鑒定專家有義務到場陳述自己主張。同樣，我國刑事訴訟法建立了鑒定人出庭作證制度，並且允許控辯雙方申請專家輔助人出庭協助對鑒定意見進行質證，在一定程度上允許引入檢驗報告。因此，從各國鑒定制度發展趨勢來看，鑒定制度通過設計各種環節，促使專業知識開示或被理解。

鑒定法律制度並沒有對算法認知起到開示作用。算法認知在鑒定過程中廣泛存在。隨著人工智能新工具、新模型不斷開發，法庭科學可實現的範圍越來越寬。^[18] 算法與法庭科學技術深度融合，DNA分型技術、人臉識別、GPS追蹤技術新型法庭科學技術層出不窮。但現有鑒定制度並沒有對算法認知起到開示功能。欲論證這個結論，筆者從如下兩個方面進行展開論述：一是，算法與鑒定工作如何進行融合；二是，當前鑒定制度並無有效機制開示算法認知。

值得說明，筆者之所以選取鑒定制度為樣本進行分析，理由如下：一方面，鑒定制度在1979年便寫入我國刑事訴訟法之中，後面經歷了1996年的微調以及2012年系統化，形成了初具規模的制度體系。^[19] 另一方面，當前專門性問題解決的格局都以鑒定制度為參照。專門性問題的解決上，我國形成了鑒定意見、專門性報告以及事故調查報告等多元格局，^[20] 審查框架以鑒定意見審查為核心，後兩者主要是參照“鑒定意見”的有關規定進行審查。^[21] 因此，對鑒定制度展開分析，足以涵蓋整個專門性問題多元格局。

司法實踐中，算法與法庭科學存在兩種融合模式：“取證複核”模式與“賦能技術”模式。“取證複核”模式是指通過算法取證後，有關機關再將分析結果拿去鑒定複核，最終在法庭上出示鑒定結果。

案例1：2018年李某來到連雲港市偷盜10條香煙，價值約8500人民幣。2020年9月，公安機關在鹽城市將李某抓獲。到案的李某與案發監控視頻容貌變化較大。面對檢察官的訊問，李某始終拒不認罪。承辦案件檢察官自行補充偵查，引導偵查機關展開人臉圖像對比工作，確定圖像的同一性。將李某與案發酒店當天監控提取到的人臉畫面放置江蘇省相關信息進行比對，結果顯示，李某是第一相似人，相似率高達92.48%。考慮到證明力的問題，又將李某多張照片以及作案前半個小時的監控視頻中人臉畫面作為檢材，送至外省專業鑒定機構進行鑒定，鑒定結果也證實圖像為同一人。^[22]

上述案例是“取證複核”模式的體現。司法機關通過人臉識別算法，鎖定犯罪嫌疑人的身份，同時出於增強證據證明力的考慮，又將犯罪嫌疑人的多張照片送至鑒定中心，讓其進行人像鑒定，保障算法識別結果的可信性。“取證複核”模式具有如下特征：第一，算法取證與鑒定具有先後順序。偵查機關通常先利用算法識別獲取初步結果，又將結果送至鑒定核實。算法識別先於鑒定。第二，鑒定檢材實際上是由算法進行選取。案例1和案例2中，鑒定對象實際上是由算法進行選取的。算法明確其為李某後，李某多張圖像被作為比對檢材送至鑒定中心，進行鑒定。第三，鑒定是

[18] 何曉丹：《從此向何處——第23屆國際法庭科學大會綜述》，載《中國司法鑒定》2024年第1期，第31頁。

[19] 參見郭華：《刑事鑒定制度修改的背景、爭議及解讀》，載《證據科學》2012年第2期，第179頁。

[20] 也有學者對專門性報告、事故調查報告的合法定種類型提出質疑。參見龍宗智：《立法原意何處尋：評2021年最高人民法院適用刑事訴訟法司法解釋》，載《中國法學》2021年4期，第247-266頁。

[21] 參見吳洪洪：《刑事訴訟專門性證據的擴張與規制》，載《法學研究》2022年第4期，第177頁。

[22] 尹賀、張梅娟、宗申：《人臉識別鎖定“零口供”盜竊慣犯》，載《檢察日報》2021年9月11日，第3版。

對算法分析結果的複核與檢驗。案例 1 中可以看出，視頻身份認定是由算法做出的關鍵判斷，而不是司法工作人員以及鑑定人員。

“賦能技術”模式是指算法賦能法庭科學技術升級迭代，技術人員利用新興技術對證據材料進行鑑定或分析。學界將法庭科學證據進化歷程，分成兩個階段：第一代法庭科學證據，例如指紋分析、彈道分析，它們的使用範圍有限，技術特征表現為經驗性與觀察性並不需要複雜的儀器去分析。第二代法庭科學證據，例如 DNA 分型、位置追蹤、生物識別技術，它們適用範圍更廣，技術直覺上無法被觀察，分析方法上也更加複雜所用儀器更加專業，技術由私營部門協助開發。^[23] 兩者核心區別在於，第一代法科證據僅單純依靠人類專業知識，第二代法科證據由源代碼引導，具有算法判斷性。^[24] 從上述演化歷程中，不難觀察到，算法與鑑定技術一步步深度融合，算法也在輔助技術人員形成認知，分析結果是人類專業知識與算法判斷的合力。

案例 2：王萬法醉酒駕車後，與朱某發生爭執，警察到達現場後對王萬法進行呼氣式酒精檢測，並將其拘留。為證明其駕駛人員是王萬法，司法鑑定所出具了視頻人像鑑定意見。鑑定意見中反映了，鑑定機關利用人臉比對系統分析了四段視頻^[25]，系統判定“對比吹氣時錄制的被鑑定目標人物的正面拍攝照片和檔案正面照片…相似度為 confidence=72.964”、“對比被鑑定目標人物和已知檔案信息相似度…相似度達到 confidence=71.675”。綜上，鑑定機構認為“對比同一人現場采集信息同檔案圖片對比結果，發現兩次比對相似度評分接近，因而判定為同一人的可能性高。故綜合分析，王萬法應當為皖 F ××××× 號小型普通客車駕駛員”。^[26]

上述案例反映了“賦能技術”模式，它具備如下特征：首先，算法與鑑定並不具備先後順序。公安機關並沒有預先對人臉圖像進行識別，而是送至鑑定機構，由鑑定機構利用算法進行識別分析。其次，檢材是由司法機關選取。從上述四段視頻看出，視頻片段選取與移送是公安司法機關完成，而不是算法判斷。最後，鑑定結果是算法與技術人員共同完成。系統對比結果呈現出“相似度 confidence=72.964”以及“相似度達到 confidence=71.675”，具有明顯的算法判斷性。最終得出的鑑定意見，則是技術人員基於算法判斷所形成的。

上述分析表明，算法通過“取證複核”模式以及“賦能技術”模式與鑑定制度融合，而在法律制度層面及時沒有對融合做出回應。首先，刑事訴訟法並未賦予算法參與資格。根據《刑訴法》第 146 條規定，鑑定主體應為具有專門知識的人。按照刑訴理論，鑑定人作為訴訟參與人，應當是自然人。^[27] 但在“取證複核”模式以及“賦能技術”模式下，算法認知作為“程序內認知決策主體”，無疑參與了司法決策，直接影響了訴訟進程與結果。^[28] 法律規範以及傳統的訴訟法理論，並

[23] Murphy, Erin., *The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence.*, 95California Law Review, p.721 (2007).

[24] 也有學者提出第三代法科證據。他認為，第三代法庭科學證據更加智能化。例如自動駕駛過程中，評估人類駕駛員的疲勞程度的系統等智能型機器證據，他們可能根據駕駛員的姿勢、面部動態來評估駕駛員是否困倦，發出並儲存警報，可能在刑事審判中作為證據使用。參見 Gless, Sabine., *AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials.*, 51Georgetown Journal of International Law, p.195 (2020).

[25] 分別是視頻《S202省道與南黎路西向-1-20180520230502-20180520231059-434015357》、視頻《南黎路與S202交口東300米南XH191-1-20180520230600-20180521004815-19250476》、視頻《南黎路與S202交口東300米南XH191-1-20180520230600-20180521004815-19250476》以及視頻《呼氣王萬法》。

[26] 參見王萬法危險駕駛案，安徽省淮北市相山區人民法院(2018)皖0603刑初336號刑事判決書。

[27] 參見張建偉：《刑事訴訟法通義》，北京大學出版社2016年版，第127頁；陳光中：《刑事訴訟法》，北京大學出版社高等教育出版社2021年版，第86頁。

[28] 參見謝澍：《刑事訴訟主體理論的揚棄與超越》，載《中國法學》2023年第3期，第273頁。

沒有給予算法認知足夠的規範空間與理論空間。其次，補充鑒定、重新鑒定可能會被虛置。補充鑒定、重新鑒定是當事人異議權的體現。“取證複核”模式下，檢材選取是由算法所決定的，由於算法認知隱而不彰，當事人無法針對檢材提出異議，最終無法通過申請啟動程序。最後，鑒定人出庭僅展示自我認知，無法展示算法認知。“取證複核”模式下，鑒定人僅能展示所送檢材與分析報告之間的專業知識，無法解釋說明檢材選取的邏輯。“技術賦能”模式下，鑒定人員依據技術可能是專門公司進行開發，鑒定人員並不知曉源代碼以及運行過程。在庭審中，也無法充分展示所依據技術的原理。簡言之，現有制度下，鑒定人出庭也難以開示算法認知。綜上分析，現有鑒定制度無法開示算法認知，算法認知隱而不彰。

3. 庭審過程中算法證據概括式審查

為應對各種犯罪，算法技術在刑事領域業已廣泛運用，誕生了諸多新型證據。例如，人臉識別技術所產生的對比報告、稅務總局利用算法分析涉案公司財務異常的報告以及行蹤軌跡對比報告等。針對實踐中產生的新興證據，理論界形成了三套話語體系對其命名，分別是：“大數據證據”^[29]、“人工智能證據”^[30]以及“算法證據”^[31]。算法是新興技術衍生證據的源頭，也是數字生活世界的核心。^[32]在證據法體系內沿用，可以保持概念的開放性以及多學科交叉的一致性。^[33]由此，筆者用“算法證據”代指通過算法對案內外大數據分析後所產生的分析報告。

算法證據在庭審中存在概括運用模式，它存在“簡化式”運用與“轉化式”運用兩種樣態。

“簡化式”審查是指司法機關工作人員不加審查，直接採納算法證據。例如，李某販毒案中，辯護律師指出經“大數據比對”認定李某時販賣毒品的人，但公訴機關並未提交大數據比對的相關證據。法院並沒有給予回應，而是籠統寫到“上述證據取證程序和證據表現形式合法，內容能相互印證，予以采信”。^[34]簡約化運用背後是法官隱藏算法證據的看法，掩飾比說理重要。^[35]同樣，這也表現出，法官責任制約下，法官面對新型科技時具有明顯的自我防衛傾向。簡化式運用中，法官以程式化說理或有意的忽略掩蓋了算法的運用。

“轉化式”審查是指司法工作人員對算法證據的審查轉化為既有法定證據種類以及證據結論真實性審查。一方面，將證據種類轉化為既有法定證據種類。以人臉識別證據為例，在法庭中，檢察機關常常將人臉識別證據轉化為“人臉對比截圖”^[36]、“人臉對比報告或材料”^[37]、“到案經

[29] 參見劉品新：《論大數據證據》，載《環球法律評論》2019年第1期，第21頁；張吉喜、孔德倫：《論刑事訴訟中的大數據證據》，載《貴州大學學報（社會科學版）》2020年第4期，第82頁；。

[30] 參見馬國洋：《人工智能證據適用的雙重風險及其規制》，載《大連理工大學學報》2023年第4期；馬國洋：《論刑事訴訟中人工智能證據的審查》，載《中國刑事法雜誌》2021年第5期，第158頁。

[31] 參見張迪：《刑事訴訟中的算法證據：概念、機理及其運用》，載《河南大學學報》2023年第3期，第36頁。張迪：《算法證據的獨立：法理反思與制度方案》，載《中國刑事法雜誌》2023年第5期，第107頁。

[32] 對數字生活世界的論述，參見[英]薩斯坎德：《算法的力量——人類如何共同生存》，李大白譯，北京日報出版社2022年版，第3頁。

[33] 對“算法證據”概念的詳細論述，參見張迪：《刑事訴訟中的算法證據：概念、機理及其運用》，載《河南大學學報》2023年第3期，第36-42頁。張迪：《算法證據的獨立：法理反思與制度方案》，載《中國刑事法雜誌》2023年第5期，第107頁。

[34] 參見李文、劉濤販賣、運輸賣毒品案，雲南省紅河哈尼族彝族自治州中級人民法院(2019)雲25刑初84號刑事判決書。

[35] 凌斌：《法官如何說理：中國經驗與普遍原理》，載《中國法學》2015年第5期，第102頁。

[36] 參見程大磊盜竊一審刑事判決書，北京市豐台區人民法院(2019)京0106刑初1434號刑事判決書。

[37] 參見任春雷盜竊一審刑事判決書，北京市豐台區人民法院(2019)京0106刑初1130號刑事判決書。

過”^[38]等證據材料形式，在法庭中出示。另一方面，法官在審查算法證據時，將證據審查轉化為證據結論的真實性審查。例如，通過人臉識別證據來分析研判人身同一性的案件中，證據合法性、關聯性常常被忽視，法官更加側重視頻中所示人像是否為犯罪嫌疑人本人。法官往往通過兩種策略進行認定：第一，借助被告人的身體特征或衣著特征結合人臉識別證據加以認定其結論真實性；^[39]第二，與其他算法證據形成證據鎖鏈來印證算法證據的真實性問題。^[40]綜上分析，算法證據的轉化式運用，僅僅簡單嫁接到既有法定證據種類之下或片面的分析算法結論的真實性，忽略了算法證據生成過程的特殊性以及內在結構的複雜性，^[41]更是掩蓋了算法證據內部運行規律。

（二）算法遮蔽式運用的風險

1. 算法遮蔽式運用導致算法內容封閉化與零散化

算法遮蔽式運用造成了算法內容的封閉化。在“弱二元主體模式”下，技術研發人員按照偵查機關要求設計相應算法，算法當中冗雜了偵查機關對刑事法律規範、刑事政策、治安管理經驗等多維知識理解。考慮到公安偵查機關的訴訟地位以及工作職能上，“弱二元主體”模式極易形成偏見，影響算法公正運行。一方面，公安偵查機關屬於控訴機關，其思維認知主要傾向於，搜集證據打擊犯罪。“弱二元主體”模式下，依賴於公安偵查機關認知設計出的算法，更偏向“重打擊，輕保護”。另一方面，公安機關既具有搜集證據、打擊犯罪的刑事偵查職能，也兼具治安維穩等行政職能。按照公安經驗與要求設計的算法可能會忽視兩者的差異，造成治安標準替代刑事偵查標準的後果。有學者也提出大數據警務核心議題便是對治安問題進行有效防控，其中既包括違反治安管理類法規和刑法問題，也包括治安法未規定的客觀存在的危害社會治安秩序的行為。^[42]上述觀點從問題發生角度強調，大數據警務需要對前端進行治理，但沒有意識到刑事偵查與違反治安管理行為的調查具有本質區別，而試圖采用相同算法內容對待。綜上所述，“弱二元主體模式”由於放棄了“警方、技術研發人員、專家、民眾”等多元主體開放參與，容易導致警方固有偏見成為算法判斷的要素，使得算法滑向“打擊犯罪”一端。^[43]

算法遮蔽式運用造成了算法內容的零散化。學界普遍達成一種共識：算法即法律。^[44]相較於立法活動的嚴格程序性以及法律統一性，各個部門甚至各個警種都具有定制算法的權力，算法的產生具有隨意性。算法產生的隨意性導致算法內容的零散化，體現在兩個方面：一方面，各個地方所面臨相同問題，由於“算法定制權”的分散，可能設置不同算法內容。另一方面，同一地方各警種之間，出於執法分工不同且都具有“算法定制權”，也可能設置不同算法內容。算法內容的零散化可能會造成如下危害：一是，各地算法內容不統一，難以形成打擊犯罪合力。以英國 HART 系統為例，該系統收集數據主要是達勒姆郡警察局的積壓案件，針對數據提供相應的算法也具有一定的局限性，不能在其他地區警局適用。^[45]我國可能在合成作戰中實現共享，達成相對統一標準，對於

[38] 參見張建輝盜竊一審刑事判決書，浙江省蒼南縣人民法院(2018)浙0327刑初1073號刑事判決書。

[39] 參見彭雄亮盜竊案，廣西壯族自治區柳州市城中區人民法院(2020)桂0202刑初280號刑事判決書。

[40] 參見潘國星盜竊罪案，湖南省武岡市人民法院(2022)湘0581刑初347號刑事判決書。

[41] 對於生成原因的論述，參見張迪：《算法證據的獨立：法理反思與制度方案》，載《中國刑事法雜誌》2023年第5期，第113頁；對結構論述，參見周慕涵：《論大數據證據的法律地位》，載《法律科學》2023年第4期，第106頁。

[42] 李金峰、陳如超：《大數據警務風險及其規制》，載《中國人民公安大學學報》2023年第5期，第127頁。

[43] 相同的觀點參見沈國琴、齊小力：《人工智能嵌入預測警務的法律風險及其預防》，載《廣西社會科學》2021年版第5期，第14頁；馬方、唐娜：《社會安全風險視域下預測性偵查的風險與多元治理》，載《蘭州學刊》2023年第11期，第97頁。

[44] [美]勞倫斯·萊斯格：《代碼2.0：網絡空間中的法律》，李旭、沈偉偉譯，清華大學出版社2009年版，第6頁。

[45] Morison, John, Adam Harkens. ,*Re-Engineering Justice: Robot Judges, Computerised Courts and (Semi) Automated*

基層民警辦理“小案”並不能直接享受紅利。^[46]二是，算法內容零散化可能會導致算法功能蠕變。所謂功能蠕變也稱之為“功能淺變”是指技術或者系統的適用逐漸擴大，直至超出最初預期的用途。^[47]例如，疫情期間某地方利用健康碼攔截上訪。零散化算法可能會異化淪為地方治理的工具。

2. 算法遮蔽式运用导致技术性正当程序缺失

數字時代下，傳統正當程序理論難以應對新興算法問題，技術性正當程序應運而生。希特倫前瞻性的提出“技術性正當程序理論”，以應對行政程序中數字技術對正當程序的沖擊。^[48]他主要圍繞“內置規範選擇、系統算法設計、系統審計軌跡以及消除聽證偏見”等四個方面展開。行政法領域內也有部分學者參照其理論框架，提出相應的理論框架：技術性正當程序理論規制框架可擴展為“四維度和八要素”；^[49]也有學者參照哈貝馬斯“交往溝通理論”，提出自動化決策三個核心要素：排除偏見、聽取意見和說明理由。^[50]還有學者指出，技術性正當程序理論實踐和理論主張可以被重構成一套理想模型，其中包括全程參與、原理公開、充分告知、有效交流、留存記錄、人工審查等六個方面。^[51]隨著算法問題與刑事司法深度綁定，刑事訴訟領域也出現了傳統正當性理論被新興技術侵蝕的問題，技術性正當程序理論也跨越了行政法的邊界，進入刑訴學者視野。有學者指出，技術正當程序是一種新的程序正義觀，核心使命在於克服傳統正當程序理論對人工智能背景下司法運作的解釋力失靈的問題，具體包括：推進司法數據公開與算法透明、推進算法可解釋性構建問責機制。^[52]也有學者圍繞“技法兼容”的思路探討刑事司法中技術性正當程序的基本內涵，其中包括以排除算法偏見貫徹透明原則為核心的裁判中立和程序公開，過程控制和權利保障為核心的程序參與和程序對等以及以算法問責為核心的程序正義的保障措施。^[53]也有學者指出刑事司法領域技術性程序正義理論。應當包括：排除偏見（司法數據和算法透明、相應風險提示）、充分參與（系統開發運用過程中共同體與利害關係人的參與）、程序對等（保障利害關係人對智能系統的質疑機會）、程序合理（說明理由、合理解釋與反對自動化裁決）。^[54]綜上分析，上述要素都可以涵蓋在程序參與、程序公開、程序中立等三個維度之下。^[55]

算法遮蔽式運用無疑會對程序參與、程序公開、程序對等產生消極影響。

Legal Decision-Making, 39Legal Studies.p.618 (1964).

- [46] 韓關峰、陳剛：《隱私計算在大數據偵查中的應用研究》，載《中國人民公安大學學報（社會科學版）》2023年第4期，第60頁。
- [47] Paul W. Grimm, Maura R. Grossman & Gordon V. Cormack, *Artificial Intelligence as Evidence*, 19 *Northwestern Journal of Technology and Intellectual Property*, p.9 (2021).
- [48] Danielle Keats Citron, *Technological Due Process*, 85 *Washington University Law Review*, p.1249 (2008).
- [49] 四維度八要素：程序參與維度、程序公開維度、程序中立維度、程序公正維度，八要素：測試評估、審計軌跡、公眾參與、理性聽證、原理公開、信息告知、限制材料與人為幹預。參見趙龍：《自動化行政的技術性正當程序規制研究》，載《法律科學》2024年第3期，第94頁。
- [50] 排除偏見包括算法的公開透明和程序的一致性；聽取意見包括算法的可解釋性；說明理由包括允許質疑事後聽證專業審計和及時糾錯。參見劉東亮：《技術性正當程序：人工智能時代程序法和算法的雙重變奏》，載《比較法學研究》2020年第5期，第73頁。
- [51] 蘇宇：《數字時代的技術性正當程序：理論檢視與制度構建》，載《法學研究》2023年第1期，第91頁。
- [52] 李訓虎：《刑事司法人工智能的包容性規制》，載《中國社會科學》2021年第2期，第56-59頁。
- [53] 餘鵬文：《現象、原理和規制：人工智能司法與刑事程序正義的融合之路》，載《天府新論》2023年第1期，第121-123頁。
- [54] 劉金松：《數字時代刑事正當程序的重構：一種技術性程序正義理論》，載《華南科技大學學報（社會科學版）》2023年第2期，第21-23頁。
- [55] 程序公正的意義過於廣泛，程序公正本身就涵蓋了程序公開、程序參與以及程序對等，因此不宜將其設立為獨立維度。

首先，算法遮蔽式運用導致當事人無法實質參與到遮蔽式運作之中，無法為自己或委托人爭取合法利益。程序參與的核心要義是“權益可能會受到刑事裁判或刑事審判結局直接影響的主體應有充分的機會並富有意義地參與法庭裁判制作過程，從而對法律裁判結果的形成發揮有效的影響和作用。”^[56]在算法遮蔽式運作過程中，當事人程序參與權常被架空。一方面，偵查階段，算法設計封閉性使得有利於辯方立場之數據與需求，難以實質納入設計考量之中；封閉式算法取證架空以立案為端口的偵查啟動控制模式，立案前便開始有關數據收集，侵犯公民隱私權。對此無接觸的侵權，當事人並沒有形成感知，加之相關證據合法性審查法律規範疏漏，當事人自然無法參與到程序之中。另一方面，審判階段對算法證據概括式審查，無法有效保證當事人對質權與異議權。換言之，對於極大影響事實認定者心證的關鍵證據，辯方無法通過有意義的質證來削弱其證明力。

其次，程序公開自然與算法遮蔽式運用相矛盾。算法的技術性黑箱，是人們難以理解輸入與輸出之間存在的技術隱層。刑事訴訟程序中隱蔽式運用，進一步隱藏了算法黑箱，使得本就難以理解的算法技術，排斥監督與審查。偵查機關啟動算法取證時，通常以偵查機關內部高層級審查為主，排斥司法機關審查批准。

最後，算法遮蔽式運用難以保障程序對等。程序對等是指控辯雙方都應在影響裁判心證方面，有著平等的機會、便利與手段。裁判也應對各方意見與證據予以平等的關注。^[57]算法遮蔽式運用反映了算法從設計、運行到形成最終裁判結論全流程中，辯方往往難以有效參與到各環節之中，並且與控方所擁有的算力並不對等，雙方處理數據實現目標結果的能力具有天壤之別。^[58]同時，算法遮蔽式運用中，法律賦予辯護方的救濟權，難以達成預設控辯平等對抗之效果。例如，“取證複核”模式下，重新鑒定與補充鑒定辯本是辯方之救濟權，但無法針對算法所選檢材提出異議，難以達成重新鑒定或補充鑒定之目的。

3. 算法遮蔽式運用可能助長司法工作人員的依賴性

算法在我國司法實踐中普遍運用。受到實踐需求以及政策影響，算法深度融合與刑事訴訟全流程。一是，在偵查階段，算法助力偵查人員發現犯罪線索和預測犯罪。^[59]二是，在檢察階段，數字檢察戰略驅使各地檢察機關形成獨立模型。利用算法碰撞，發現案件線索。例如，部分檢察機關通過建模中心，依托“算子”進行大數據碰撞，發現 80 多條販毒刑事犯罪立案監督線索。^[60]三是，在審判階段，算法證據層出不窮。有學者指出，算法證據極易影響法官心證形成，構成庭審中的“數據獨裁”。^[61]可見，算法的顯著優勢就是在案件中，高效輔助數據整合，為訴訟提供必要的線索與證據材料。

算法高效組織數據，遠超人工組織梳理，可能會導致司法人員對事實查明過程出現算法依賴，偏向於算法支撐的證據材料。^[62]同時，原本由司法人員依據法律程序做出判斷過程，變轉為算法依據數據庫得出分析結論的過程。^[63]

[56] 陳瑞華：《刑事訴訟的法理》，商務印書館2021年版，第228頁。

[57] 同上註，第232頁。

[58] 鄭曦：《數字時代刑事證據開示制度之重塑》，載《華東政法大學學報》2023年第4期，第41頁。

[59] 韓關峰、陳剛：《隱私計算在大數據偵查中的應用研究》，載《中國人民公安大學學報（社會科學版）》2023年第4期，第60頁。

[60] 賈宇：《論數字檢察》，載《中國法學》2023年第1期，第15頁。

[61] 程龍：《論大數據證據質證的形式化及其實質化路徑》，載《政治與法律》2022年第5期，第100頁。

[62] 魏俊斌：《大數據偵查的風險防控與“意識”匡正》，載《大連理工大學學報（社會科學版）》2024年第4期，第93頁。

[63] 資金分析報告的適用。

對此情況，法律未來主義者樂觀其成。在他們眼中，人類社會制度構造的法律制度只不過是一種算法而已。^[64]對於具備高度算力的算法模型而言，人類學習能力本身嚴重滯後，功用性遠不及算法模型。^[65]

從上述實踐與理論發展中，我們不免看到一絲隱憂：司法人員主體性的退卻。算法的遮蔽式運用進一步隱藏了這種退卻。換言之，法律制度層面並沒有提供一個充分的開示空間，更助長司法工作人員的依賴性。

三、構建刑事算法開示制度的必要性

針對算法的遮蔽式運用以及風險，構建刑事算法開示制度具有極端重要性。

（一）算法開示制度有助於促進統一完善的標準形成

算法開示可以完善算法內容設計，盡可能避免偏見引入。算法偏見可以參照計算機偏見，按照來源可以被分成：先行存在偏見、技術性偏見和突發性偏見。^[66]先行存在偏見是指算法設計者載編程時可能將自身偏見、社會風氣、制度體制嵌入到算法之內。技術性算法偏見是指主要由於技術本身以及計算能力或者是系統某些限制造成的。突發性算法偏見是指設計算法之時未能預見，由於新的社會形勢變化，導致算法不相關。技術性算法偏見無法通過法律制度安排給予解決。筆者重點分析，通過算法開示解決先行存在偏見與突發性偏見。對於先行存在偏見而言，算法開示將先行存在偏見外顯化，通過社會或法制固有的倫理加以引導或者消除。^[67]例如，公安機關設計相關算法，可能存在“重打擊，輕保護”的先行偏見，通過算法開示將偏見外顯化，在善法引導下，修正相關算法。對突發性偏見而言，當刑事法評價要素突發性轉變，算法設計存在滯後性。算法開示在訴訟制度內部提供了必要開放路徑，已達到信息反饋與糾偏作用。以酒駕為例，2024年3月1日實施的新國標，飲酒駕駛的上限從50mg/100mL降低到了20mg/100mL，同時，新國標將血液中乙醇含量與呼吸中乙醇含量之間的換算系數從2100調整為2300。隨著新國標施行，酒精分析儀中內置算法也應做出相應調整。通過算法開示，可以顯現出算法是否符合新國標，對儀器的適用性提供信息反饋。

算法開示可以促進多元主體參與，打破算法孤島。一方面，算法開示為多元主體參與提供制度空間。對律師而言，通過閱卷算法，來判斷算法是否被誤用、算法是否存在不可靠性；對於檢察人員而言，破除封閉性將促進有效監督。^[68]對於審判人員，可以強化對算法取證行為的實質性審查。另一方面，算法開示可以保證多元參與設計論證，避免嵌入偏見認知。以美國司法實踐為例，美國賓州量刑委員會圍繞量刑算法召開了19次聽證會，征求公眾對量刑算法公式的意見與反饋，最終排除了涉嫌種族歧視與階級歧視的郵政編碼等因素。^[69]

算法開示可以防止危害成本外化，防止社會不公蔓延。算法時代下，由於感知設備安裝與佩戴，一舉一動都留下了電子痕跡，變成了可供處理、分析和利用的數據，每個人都具有“被強加給

[64] 蔣舸：《作為算法的法律》，載《清華法學》2019年第1期。

[65] 余成峰：《法律的“死亡”：人工智能時代的法律功能危機》，載《華東政法大學學報》2018年第2期。

[66] 劉友華：《算法偏見及其規制路徑研究》，載《法學雜誌》2019年第6期，第59頁。

[67] 楊慶峰：《數據偏見是否可以消除》，載《自然辯證法研究》2019年第8期，第112頁。

[68] 李金峰、陳如超：《大數據警務風險及其規制》，載《中國人民公安大學學報（社會科學版）》2023年第5期，第135頁。

[69] 42 Pa. Cons. Stat. § 2155(a)(1) (2023).

個人的量化身份”。^[70] 算法通過對公民數字身份重構，創造了新的階層和類別。算法決策重複性使得某一領域遭遇算法歧視的主體被結構性鎖定，導致未來發展系統性不利。在算法普遍存在情況下，某一領域被打上標記，可能會造成算法系統之間的連鎖反應。算法世界也會與現實世界發生共振，加劇現實社會中的問題。例如，ProPublica 對 Campus 審查後，發現該軟件將非裔美國人標記為高風險但實際未犯罪可能性是白人群體的兩倍。^[71] 這也會加劇現實世界中美國的社會割裂。算法開示可以將算法進行合理公開，吸引公眾合理參與，協助確定算法影響的消極程度，為監管與幹預提供必要的指引與制度基礎。

（二）算法開示制度奠定刑事技術性正當程序基礎

算法開示是程序參與的基礎。1975 年，Thibaut 與 Walker 研究提出了程序正義的“控制理論”，即相對於從第三方獲取所欲求的判決結果，人們更在意自身對決定產生了何種程度的控制性影響。這種影響構成了感知程序正義是否公正的標準。^[72] 換言之，有效的程序參與才能保證程序可感知的公正。算法開示是程序參與的基礎：一方面，算法開示可以實現知識公開。依照福柯知識-權力論，科學話語的形成深受權力的內在推動，兩者相輔相成，處於持續動態互動之中。^[73] 在算法遮蔽式運行下，控辯雙方算力不對等，本質上是一種知識不對等。算法開示盡可能提供給辯護方關於算法足夠知識，使得控辯雙方相同知識參與到訴訟競技之中。另一方面，算法開示可以吸納多元主體參與。算法治理具有普遍性，關切每個公民自身權益，需要多元主體協同發力，構築良性運行生態。算法技術黑箱加之司法運用過程中被相關機構有意識遮蔽，導致算法不為公眾所知。算法開示後，有利於吸納多元主體協同參與。

算法開示是程序公開的應有之義。程序公開是司法程序的開始與推進應當對外公開而不能秘密進行。^[74] 程序不公開造成後果便是權力缺乏監督。算法遮蔽式運用造成司法機關外部監督乏力以及內部審查闕如的弊病，可以通過算法開示予以紓解。從司法機關外部監督的角度，算法開示可以保證當事人、新聞媒體以及公眾對算法原理與運用，享有知情權，規範算法的應用。例如，從美國算法運用來看，新聞媒體在其中扮演了“第四權力”的重要角色，諸多關鍵技術運用錯誤，都是由媒體以及社會團體予以披露，這也促進了參議院諸多提案誕生。從司法機關內部審查角度，偵查機關的內部審批程序存在封閉性、自我賦權等弊端。當務之急，乃是在構建司法審查機制，以權力制約權力。構建司法審查機制關鍵在於算法取證需要對審查人員開示，說服審查人員簽發令狀。在此基礎上，可以說算法開示是未來構建算法取證的司法審查措施的關鍵環節。

算法開示有助於保障程序對等。算法遮蔽式運用中，控辯雙方並不對等。一方面，算法時代下，諸多訴訟權利收效甚微。例如，大數據證據質證存在形式化的問題。另一方面，從證據偏在以及算法獨裁角度而言，算法取得的證據被控方用於指控，加之其具備“科學性”的表象，事實認定者心證天平無疑會偏向控方。由此，算法開示如下兩個方面保障程序對等：第一，算法開示可以激活配套訴訟制度，協同保障被告人利益。例如，經算法開示後，可以明確訴訟爭點以及可以對算法提出相應質疑，有效踐行被告人質證權。第二，算法開示可以影響法官心證。算法開示所得到的證據材料，可用於對算法的彈劾，動搖法官心證。同時，算法開示過程中，法官也可以得到相關信

[70] 鄭戈：《算法的法律與法律的算法》，載《中國法律評論》2018年第2期，第67頁。

[71] Melissa Hamilton, *The Biased Algorithm: Evidence of Disparate Impact on Hispanics*, 56 American Criminal Law Review, p.1553.(2019).

[72] John Thibaut & Laurens Walker, *A Theory of Procedure*, 66 California Law Review, p.541.(1978).

[73] 施冠男、劉同舫：《科學話語何以主導文化體系——福柯“知識-權力”結構的理論效應》，載《自然辯證法通訊》2024年第8期，第102-109頁。

[74] 江必新、程琥：《司法程序公開研究》，載《法律適用》2014年第1期，第27頁。

息，形成對算法的自我認知。

（三）算法開示制度可以促進算法問責

問責制度是國家治理必備要素。習總書記在十九屆中央政治局民主生活會上也曾強調“要強化責任意識，知責於心、擔責於身、履責於行”。當前，算法遮蔽式運用造成了傳統問責變得困難重重。具體而言，算法的深度參與導致了訴訟流程中做出決策的主體從單一化邁向了多元化與複雜化，決策主體不再明晰，導致難以有效問責。^[75]由此，衍生出“算法問責”這一概念。據學者考察，算法問責制的完整定義來自於萬維網基金會，即“在算法系統中確保損害可以被評估、控制和補救”。^[76]

算法開示可以從識別責任主體以及克服問責障礙方面，促進算法問責。一方面，算法開示可以有助於識別責任主體。識別責任主體，即向誰問責，算法問責的先決條件。決策主體的多元化以及算法遮蔽式運行導致了算法問責難以實現。算法開示有助於準確劃分算法與決策人員之間的責任。根據司法工作人員與算法之間的關係，可以將算法決策分成自主決策與輔助決策，也有學者稱之為“決策替代系統”與“決策支持系統”。^[77]自主決策是指算法在決策中，不需要人的參與，進行得出結論。例如，“取證複核”模式下算法選取檢材的過程。輔助決策是指算法扮演輔助性角色，決策者仍有較大的裁量權。例如，“賦能技術”模式下，算法輔助鑒定人員形成決策。通過算法開示可以認識算法認知介入的邊界在哪，從而有效識別責任主體。另一方面，有論者分析指出算法問責難以實現，原因在於透明度困境導致公眾對司法工作人員喪失信任；使用者難以控制算法，算法迭代可以按照預想方向改進；決策者向公眾回應時無法解釋說理。^[78]概括言之，算法對司法工作人員以及對公眾均不透明，影響公眾知情權和參與權。算法開示奠定了技術性正當程序基礎，保障公眾知情權與參與權。另外，從域外實踐來看，算法問責以算法開示為核心。根據紐約市算法問責法、華盛頓州算法問責聽證會以及美國國會議員提出問責法案，美國算法問責制度的核心是算法使用者或控制者有義務報告並證明算法系統設計和決策應用的合理性。^[79]

四、刑事算法開示制度構建的邏輯

算法開示有兩條清晰的法理脈絡支撐：依托於職權、服務於訴訟效率的權力邏輯；旨在保障訴訟參與、接受公正審判的權利邏輯。前者保證了算法開示制度的可行性，即算法開示制度實然運行之保障；後者明確了算法開示制度的正當性，是應然理想之圖景，是算法開示未來發展的方向。這兩條法理脈絡是算法開示程序對國家權力與公民權利平衡的深層次安排。

（一）權力行使邏輯：依托職權、服務於訴訟效率

算法開示的權力行使邏輯，指的是司法機關在開示相關算法時，將其視為一種依職權行使的追訴性活動看待，而非對被迫訴人的權利保障活動。換言之，司法機關將算法開示作為換取被迫訴人認罪的一種策略，目的在於使被迫訴人意識到，司法工作人員已經掌握其充分犯罪證據，從而使其放棄抵抗。這並非一種對算法開示惡意揣測，而是針對已發生現實的概括。

[75] 參見李訓虎：《刑事司法人工智能的包容性規制》，載《中國社會科學》2021年第2期，第42-62頁。

[76] 郭春鎮、勇琪：《算法的程序正義》，載《中國政法大學學報》2023年第1期，第166頁。

[77] 高童非：《論自動化行政中“算法卸責”》，載《電子政務》2023年第2期，第118頁。

[78] 高童非：《論自動化行政中“算法卸責”》，載《電子政務》2023年第2期，第121-124頁。

[79] 張欣：《從算法危機到算法信任：算法治理的多元方案和本土化路徑》，載《華東政法大學學報》2019年第6期，第23頁。

值得說明的是，雖然筆者提及，算法開示並不同於算法證據開示，而對於兩者內在邏輯卻是相似的。實踐中已經出現，司法機關向被迫訴人展示算法取得的證據等具有開示性質的訴訟活動。有理由相信，相同刑事訴訟體制下，兩者會具有一致的邏輯。

權力行使邏輯蘊含在最高檢頒布的相應指導案例之中。指導性案例承載著統一司法的作用，它們由最高司法機關事先遴選公布，是貫徹最高司法機關意志的工具。^[80]同時，也是我們觀察法律制度運行邏輯的重要切口。大部分檢察機關通過對算法取得的證據進行開示，讓犯罪嫌疑人放棄僥倖心理，簽署認罪認罰協議，達成考察指標。例如，廖某甲等4人偷渡國境案中，廖某甲等4人商議，從雲南出境，結夥偷渡緬甸。最終，4人成功偷渡後，因工作環境與預期差距較大以及新冠肺炎疫情嚴重，4個被告人決定回國。他們從雲南入境時，被當地公安查獲並處以行政處罰。4人中廖某乙與鄭某乙辯稱單獨偷渡、否認結夥偷渡。檢察機關建議偵查機關借助省公安廳大數據平台，調取到4人留下同一航班、同一高鐵、同一住宿記錄、手機基站分布情況交通出行數據。在審查起訴階段，檢察機關通過向始終不認罪的廖某乙、鄭某乙開示交通出行數據、闡明認罪認罰制度，促使兩人在證據面前選擇認罪認罰。^[81]又如，M、N全國連鎖飲品企業運營經理常某某職務侵占案中，司法工作人員通過對犯罪嫌疑人支付寶、微信收款大數據篩選司法審計可信性和證明力，為開展犯罪嫌疑人認罪認罰打好根基。^[82]

從上述案件中，反映了算法開示的權力行使邏輯的脈絡。一方面，職權邏輯占據主導地位。證據開示的啟動、內容以及時間節點都是檢察機關依照職權確定。同時，從指導案例的表述中，檢察機關意圖並非讓當事人獲知其通過算法取得的證據信息，而是在於向被迫訴人施加壓力或保障追訴的正確性。另一方面，權力行使邏輯有助於提升訴訟效率。相較於形成完整的證據鏈條，並經過實質性庭審方可對被迫訴人定罪，認罪認罰制度大大提升了訴訟效率。檢察機關通過展示算法取得的證據，來證明自己有十足的把握在後續訴訟流程中占據優勢。這對被迫訴人形成一種威壓，來迫使他們簽署認罪認罰協議，大大提高了訴訟效率。

算法開示權力行使邏輯的形成有如下原因：首先，算法證據的印證式審查盛行。當前，法官審查算法證據大都依靠印證式審查，即依靠證據所包含信息內容的同一性以及指向的同一性進行審查，忽視對單個證據的三性審查。^[83]檢察機關在審前開示證據，獲取犯罪嫌疑人認罪口供，以此取得相關證據，與算法證據形成印證。這種做法迎合了庭審中印證式審查，保證算法證據被法庭採納，將案子辦成鐵案。其次，認罪認罰為權力行使邏輯運作提供了制度空間。上述案例以及案例1中，不難看出，檢察機關開示證據後，銜接認罪認罰程序。單純獲取被告人口供，後續仍存在庭審翻供的隱憂。相較於庭審中不確定性，檢察機關更傾向於審前攻克被迫訴人獲取口供，甚至促成被迫訴人認罪認罰。倘若適用認罪認罰程序，案件在庭審時僅需審查認罪認罰的自願性、真實性、合法性，極大程度規避了證據審查以及翻供的可能性。因此，認罪認罰為開示制度提供了必要制度空間。最後，由於當前算法本身含有“打擊犯罪”的追訴傾向，它與檢察機關的追訴職權協同共振，形塑了算法的權力行使邏輯。

（二）權利保障邏輯：保障訴訟參與、接受公正審判權

算法開示的權利保障邏輯，指的是司法機關在適用算法開示時、立法機關設置算法開示時，將

[80] 參見秦宗文：《案例指導制度的特色、難題與前景》，載《法制與社會發展》2012年第1期，第100頁。

[81] 參見《最高人民檢察院依法懲治妨害國（邊）境管理犯罪典型案例》，載中華人民共和國最高人民檢察院網站：https://www.spp.gov.cn/xwfbh/wsfbt/202207/t20220707_562224.shtml#2，2024年7月19日訪問。

[82] 最高人民檢察院網上發布廳：《檢察機關依法辦理民營企業職務侵占犯罪典型案例》，網址：https://www.spp.gov.cn/xwfbh/wsfbt/202205/t20220505_556156.shtml#1，2024年7月19日訪問。

[83] 龍宗智：《刑事印證證明新探》，載《法學研究》2017年第2期，第152頁。

算法開示的制度作為一種保障被追訴人權利。迥異於權力行使邏輯對現實的概括，權利保障邏輯為構建算法開示提供了正當性證成。制度的正當性是指特定曆史條件下，法律制度以普遍的價值體系為標準的自我價值判斷和證成。現代訴訟體制中，保障人權既占有絕對道德地位，又是未來訴訟發展方向。因此，算法開示的權利保障邏輯具有鮮明的應然屬性。

權利保障邏輯蘊含於保障訴訟參與、接受公正審判權之中。一方面，權利保障邏輯在於保障以犯罪嫌疑為核心的多元主體參與。季衛東教授曾指出“強調正義、道德以及情理的另一角度、另一任務，會要求判斷者、決定者必須考慮不同方面甚至對立的規範。”算法開示的功能在於為遮蔽式運作之角度開一扇制度之窗。因此，有了算法開示制度，通過算法開示等確定性制度空間，多元主體可以參與協商。在算法設計環節，可以更廣泛的聽取不同主體的意見；在檢察機關主導的認罪認罰環節，可以保障犯罪嫌疑人的知情權以及認罪認罰自願性等。另一方面，算法開示在於保障被告人的公正審判權。公正審判權可以細分為對質權與正當程序權。^[84]第一，算法開示制度對對質權具有保障作用。隨著算法證據頻繁出現在法庭之上，法院需要應對算法封閉與對質權的張力。個案基礎上，美國法院逐步承認算法開示對被告行使對質權至關重要。例如：Underdahl 案中，明尼蘇達州最高法院指出“如果被告提供證據表明源代碼分析可能會發現可能質疑酒精測試儀可靠性的缺陷。”^[85]Wakefield 案中，檢方使用 TrueAllele 分析 DNA，被告人指出拒絕他訪問軟件源代碼，侵犯了他的對質權。上訴法院認為被告人觀點新穎，並且裁定 TrueAllele 本質上是證詞性陳述，因此受到對質權條款約束。^[86]第二，算法開示也對正當程序條款有所影響。理論上，技術性正當權利前文已有提及，不再贅述。已有域外法院認識到，不公開算法，將會損害被告人正當權利。^[87]

權利保障邏輯本身正當性源自於三個訴訟基本理論：一是訴訟關照理論。訴訟關照是指司法機關有義務對被追訴人行使訴訟權利給予必要關照，有義務協助被追訴人充分行使其訴訟權利。^[88]權利邏輯的落腳點與訴訟關照理論理念相契合，核心都在於保障當事人的訴訟權利。二是平等武裝理論。平等武裝理論要求立法給予訴訟雙方平等的訴訟權利和攻防手段。^[89]由於算力差異以及算法證據偏在，大部分證據都掌握在控方手中。算法開示制度促進了控辯雙方信息交換，確保雙方處於大致對等水平競爭或協商。三是訴訟參與理論。與權力邏輯相對，權利邏輯中被追訴人本身不再成為辦案機關獲取口供的客體，而是被視為獨立的權利個體。權利邏輯下，被追訴人被賦予充分的權利，對訴訟結果產生影響。

權利保障邏輯在應然層面上高度契合訴訟法理，它的理論定位具有原則性與指導性，內涵具有開放性與包容性。在面對追訴犯罪需求和權力行使邏輯時，算法開示制度應當堅持權利保障優先。因為算法開示制度的定位乃是控辯雙方力量不對等、控方追訴犯罪力量過強的情形下提升辯方力量、促進雙方對抗平等的平衡器。^[90]如若權利邏輯喪失其優先地位，將折損制度本身正當性。

[84] Steven M. Bellovin, Matt Blaze, Susan Landau & Brian Owsley, *Seeking the Source: Criminal Defendants' Constitutional Right to Source Code*, 17 The Ohio State Technology Law Journal, p.1 (2021).

[85] State v. Underdahl, 767 N.W.2d 677, 685 (Minn. 2009).

[86] People v. Wakefield, 107 N.Y.S.3d 487, 495 (App. Div. 2019).

[87] State v. Kummer, No. A08-0533 (Minn. Ct. App. Oct. 7, 2008)

[88] [德]約阿希姆·赫爾曼：《〈德國刑事訴訟法典〉中譯本引言》，李昌珂譯，中國政法大學出版社1995年版，第12頁。

[89] 冀祥德：《控辯平等之現代內涵解讀》，載《政法論壇》2007年第6期，第90頁。

[90] 鄭曦：《數字時代刑事證據開示制度之重塑》，載《華東政法大學學報》2023年第4期，第41頁。

五、刑事算法開示制度構建的障礙

算法開示制度作為一種新興制度，構建時注定會與既有制度發生抵牾。明確可能存在的障礙，透視障礙背後的利益權衡，有助於構建更融貫於法律體系的制度。

（一）算法開示制度與商業、國家秘密保護制度的沖突

算法開示制度與商業秘密保護制度存在沖突。域外程序規則中確立了商業秘密保護制度。美國《聯邦證據規則》對特免權內容保持高度開放，各州承認特免權差異較大，但《聯邦證據規則（建議稿）》將商業秘密納入到最初規定特免權之中，足以顯示其重要性。加州證據規則中也對商業秘密的保護令程序、閉庭程序以及密封程序做出規定。^[91] 商業秘密保護制度背後價值在於保護開發者投資利益以及維護行業創新。算法開示制度與商業秘密的封閉具有天然的張力。在美國司法實踐中，算法開發者常常以商業秘密為由，拒絕算法開示。在實踐中，可以分成兩種情況：一是，美國法官傾向將算法認定為商業秘密。在 *Chubbs* 案中，被告要求開示 DNA 測試軟件的源代碼，但軟件公司以商業秘密為由拒絕，*Chubbs* 對此上訴，上訴法院駁回了被告開示請求。^[92] 二是，即使美國法官同意披露算法，但迫於軟件開發公司壓力，專家不敢去查看相關算法。在 *State v. Peters* 案中，法院批准專家證人簽署保密協議，審查源代碼，但是專家證人拒絕了，理由是過於嚴苛的保密協議要求以及生產廠商的脅迫。^[93] 同樣，我國《刑事訴訟法》第 54 條也可能對算法開示制度構成規範上的障礙。根據我國《刑事訴訟法》第 54 條，對涉及國家秘密、商業秘密、個人隱私的證據，應當保密。該條款將商業秘密納入到刑事訴訟保密事項之中。可見，商業秘密保護制度形成了一種“法律黑箱”，與旨在打破“黑箱”的算法開示制度產生了沖突。

算法開示制度與國家秘密保護制度也可能存在沖突。根據我國《刑事訴訟法》第 54 條，該條款同樣將國家秘密納入刑事訴訟的保密事項之中。但何為“國家秘密”，刑事訴訟法中語焉不詳。此處可援引《中華人民共和國保守國家秘密法》，對“國家秘密”這一概念進行體系性解釋。《保密法》第 13 條明確規定了七項國家秘密。其中，第 6 項規定追查刑事犯罪中的秘密事項屬於國家秘密。國家秘密保護制度背後價值在於保護國家安全。在刑事訴訟中，算法被廣泛用於追查犯罪事實，因此算法極有可能被納入國家秘密之範疇，從而拒絕被開示。

（二）算法複雜性導致無效開示

代碼複雜性與專業性可能導致無效開示。算法由代碼實現，代碼本身具有複雜性與專業性，除有專門知識的人外，其餘人幾乎無法完全理解。學界也有一種主張，認為只有公開計算機程序的源代碼，以此保障被告人憲法權利。^[94] 上述主張存在以下風險，可能導致算法開示制度流於形式：第一，被告人難以負擔辯護成本。代碼內容需要專業技術人員審核，律師團隊聘請相關技術人員，會增加辯護成本，辯護費用也水漲船高。刑事算法或大數據技術客體不僅面向重案要案，更多的是小案輕罪。例如，人臉識別算法多用於盜竊、搶奪、詐騙等案件。犯罪嫌疑人沒有足夠的財力支付高昂的辯護費用。第二，專家難以理解代碼。實踐中的算法公開印證了這一點，*Reddit* 對一部分專家

[91] 王燃、孫藝桐：《人身危險性評估的算法治理——從算法透明與商業秘密沖突展開》，載《上海政法學院學報》2023年第3期，第104頁。

[92] *People v. Superior Court (Martell Chubbs)*, B258569 (Cal. Ct. App. Jan. 9, 2015).

[93] See *State v. Peters* 2011 MT 274.

[94] Steven M. Bellovin, Matt Blaze, Susan Landau & Brian Owsley, *Seeking the Source: Criminal Defendants' Constitutional Right to Source Code*, 17 *The Ohio State Technology Law Journal* p.1-74 (2021).

公開過網站排名算法，而專家卻對算法到底如何運行存在分歧。對於算法運作機制，專家也難以理解。^[95] 經筆者親身經歷與訪談，技術人員合作大都依靠注釋來理解代碼內容。^[96] 部分程序員並不習慣代碼後面加入注釋，這也造成了代碼的可讀性缺陷。第三，專家難以解讀出偏見。技術人員通過規範文件來指導源代碼編寫、測試，規範中偏見極有可能經技術人員轉譯為帶有偏見的代碼，技術人員難以測試出規範中的偏見。

算法代碼冗長可能導致開示形式化。現有軟件實現功能複雜，開發研發團隊規模化，形成的代碼行數冗長。倘若將冗長算法全部公開，可能會導致算法開示程序淪為形式，無助於保障犯罪嫌疑人的權利。一方面，全部公開可能會拖延訴訟效率。一個企業或網站的算法由數十成百甚至上千的工程師協作完成。倘若全部公開，任由辯方審查，辯方也難以在規定的時限內完成審查。冗長的算法審查將耗費大量的時間、精力，反而可能會混淆爭點，幹擾後續訴訟程序。同時，部分算法錯誤具有間隔性，往往可能需要重複測試多次，才能出現一次。如航空飛船啟動算法可能測試 67 次中會出現一次錯誤。^[97] 另一方面，控方可能會進行“代碼傾倒”。“代碼傾倒”是指掌握“代碼”一方，將冗長代碼全部給予對方，對方無能力或無法在規定時限內審查全部代碼。根本上，架空了算法開示程序，使其變為一種走過場的形式。有學者研究指出，面對大量的信息披露，人們根本無法做到全部閱讀與自己相關的信息。^[98]

機器學習算法可能導致無效開示。機器學習算法不同於傳統算法，它通過對標記的數據集進行學習，不斷優化迭代而產生的，並非按照預設代碼運行。例如，上線不足 24 小時，微軟智能機器人 Tay 被訓練出法西斯思想。僅對機器學習的源代碼開示，並不足以揭秘其黑箱。源代碼本身對審閱者並無意義，因為源代碼只開示機器學習方法，而不是數據驅動的決策規則。^[99] 針對這一困境，有學者提出對於機器學習算法除披露源代碼之外，還應披露訓練數據。^[100] 此觀點將導致“數據傾倒”的風險。所謂“數據傾倒”是指掌握數據一方向另一方無差別、不加區分的傾倒海量數據的行為。^[101] “數據傾倒”不僅會拖延訴訟效率，而且海量訓練數據不一定具有證據法意義上的關聯性。以開示數據為名行數據傾倒之實，對辯方的辯護行為造成了困難和障礙。^[102]

（三）傳統開示制度對算法開示的阻礙

從我國現有開示制度的傳統維度出發，探求算法開示阻礙的來源，部分是由於傳統開示制度的固有問題，形成了對算法開示制度的阻礙。

我國開示制度異化為閱卷權。從功能上講，閱卷權與證據開示制度都是為了保障被迫訴方的知情權。學界普遍認為，閱卷權與證據開示制度並非相同的規則，並不能相互替代。^[103] 依托於卷

[95] 參見丁曉東：《論算法的法律規制》，載《中國社會科學》2020年第12期，第143頁。

[96] 代碼後面往往附有以“/*”起始的注釋，注釋在代碼運行中，不承擔任何功能。注釋功能在於描述編寫者編寫代碼的思路或代碼欲達成的目的。注釋只是為了提高可讀性，不會被計算機編譯。即使對於經驗豐富的程序員來說，注釋也比代碼更容易閱讀。

[97] Steven M. Bellovin, Matt Blaze, Susan Landau & Brian Owsley, *Seeking the Source: Criminal Defendants' Constitutional Right to Source Code*, 17 The Ohio State Technology Law Journal p.1-74 (2021).

[98] Cornaccia, Danielle Y., *More Than You Wanted to Know: The Failure of Mandated Disclosure*, 52 Osgoode Hall Law Journal p.673, p.673-675 (2015).

[99] Kroll, Joshua A., et al., *Accountable Algorithms*, 165 University of Pennsylvania Law Review, p.633, p.633-706 (2017).

[100] Steven M. Bellovin, Matt Blaze, Susan Landau & Brian Owsley, *Seeking the Source: Criminal Defendants' Constitutional Right to Source Code*, 17 The Ohio State Technology Law Journal p.1-74(2021)

[101] 參見程龍：《論大數據證據質證的形式化及其實質化路徑》，載《政治與法律》2022年第5期，第100頁。

[102] 鄭曦：《數字時代刑事證據開示制度之重塑》，載《華東政法大學學報》2023年第4期，第41頁。

[103] 參見劉甜甜：《認罪認罰從寬案件中的證據開示制度研究》，載《中國政法大學學報》2021年第5期，第258

宗移送主義形成的“閱卷權”制度，會對算法開示形成如下阻礙：第一，部分證據材料不會加入到案卷之中。辯方閱卷的範圍逐漸擴張。1996年《刑訴法》規定辯護律師在起訴階段享有閱卷權，僅限於本案訴訟文書和技術性鑑定材料，2012年《刑訴法》將閱卷範圍擴張到“案卷材料”，涵蓋了所有案卷卷宗（副卷除外），2018年《刑訴法》也保留了相關規定。而在司法實踐中，部分通過算法得到的證據材料，並不會被放入卷宗，尤其是預測性算法所得到的分析報告。控方往往將分析報告得到的，能夠直接證明案件事實的證據材料放入卷宗之中。辯護方自然無法從卷宗中獲知算法以及算法所得出的證據材料。第二，被告人並不享有閱卷權。根據我國《刑訴法》第40條規定，辯護律師具有閱卷權。第39條規定，辯護律師可以向被告人、犯罪嫌疑人核實有關證據。換言之，被告人並不享有如同辯護人一般的閱卷權。若秉持上述思路，僅將閱卷權視作辯護人的“固有權”，^[104]無疑會限縮被告人自我辯護之權利，也會影響到對算法證據的質證。第三，閱卷階段並未覆蓋到偵查。究因於偵查不公開原則，我國《刑訴法》第40條規定，辯護人自檢察機關對案件審查起訴之日起，方可閱卷。這一規定可能固化算法在偵查階段的“雙重黑箱”，可能影響辯護律師的辯護效果。

作為證據開示制度基礎的布雷迪規則也飽受爭議。1963年美國聯邦最高法院針對布雷迪案件，做出裁決，認為辯方提出請求的情況下，如果某證據對定罪或量刑有實質性（material）作用，那麼，控方隱瞞（suppression）這個對被告人有利的證據就是違反正當程序的，不論控方是出於善意還是惡意。^[105]後續通過對布雷迪規則完善和發展，演變為現有開示制度。基於此，我國部分學者主張，以布雷迪規則作為借鑒樣本或理論基礎，構建數字證據或大數據證據開示制度。^[106]從域外司法實踐中，布雷迪規則難以為算法開示或算法證據開示制度提供規範依據。第一，檢方未掌握源代碼，難以構成“隱瞞”（suppression）要件。City of Fargo v. Levine案中，被告人提出依照布雷迪規則，開示酒精測試儀Intoxilyzer的源代碼。北達科他州最高法院駁回此項動議，並解釋道，檢察官既不擁有也不控制Intoxilyzer的源代碼，並不構成布雷迪規則侵犯。^[107]法院邏輯在於，檢察官並未掌握源代碼，並不符合布雷迪規則中隱瞞（suppression）的要件。第二，辯護人難以對“實質性”要件證成。State v. West案件中，辯護律師要求開示源代碼，但上訴法院認為被告律師未能證明源代碼是“實質性”的，僅是簡單假定：源代碼可以用來表明，制造商偏向提供用以定罪的證據。上訴法院進一步論證到“布雷迪規則並不支持被告獲取實質性不明的證據，來測試這是否有助於或有損於案件”。^[108]由此，法院審判邏輯為辯方設計出兩難境地：辯方需要證明源代碼對定罪量刑具有實質性才能進行開示，然而，只有開示過後才知道，源代碼是否對定罪量刑重要。^[109]第三，對“有利的證據”要件的判斷依賴於法院裁量。State v. Marino案中，辯護方依照布雷迪規則申請數據與源代碼的動議，初審法院批准了控方要想辯護方律師提供所擁有的、由intoximometer生成的“所有下載或未下載的數據”，但並沒有批准源代碼。法院認為，辯方僅推測源代碼可能會影響定罪量刑。根據布雷迪規則，檢方並不需要向辯護方提交所有文件，只需要開示對辯方有利的證據

頁。程龍：《論大數據證據質證的形式化及其實質化路徑》，載《政治與法律》2022年第5期，第104頁。

[104] 參見陳學權：《論被迫訴人本人的閱卷權》，載《法商研究》2019年第4期，第103頁；陳瑞華：《論被告人的閱卷權》，載《當代法學》2013年第3期，第127頁。

[105] Brady v. Maryland, 373 U.S. 83, 87(1963).

[106] 裴煒：《論刑事數字辯護：以有效辯護為視角》，載《法治現代化研究》2020年第4期，第126頁。

[107] City of Fargo v. Levine, 747 N.W.2d 130, 2008 ND 64

[108] State v. West, 250 Or.App. 196, 279 P.3d 354

[109] Steven M. Bellovin, Matt Blaze, Susan Landau & Brian Owsley, *Seeking the Source: Criminal Defendants' Constitutional Right to Source Code*, 17 The Ohio State Technology Law Journal p.1 (2021).

即可。規則並沒有明細規定，由誰決定哪些被隱瞞的信息是有利於被告人且違反正當程序，只能交由法官進行自我裁量。這可能導致辯護人難以獲取想要的證據材料，減損辯護效果。

六、刑事算法開示制度構建的路徑

算法對各國刑事司法體制構成激烈沖擊，此種革命性的力量正在重構重塑法治運行的時空場景、功能作用、救濟途徑。從規範到實踐中，算法一直處於遮蔽式運用，應當留給刑事司法體制足夠的開放空間，圍繞算法技術，在權力行使與權利保障之間尋求微妙平衡。為此，有必要在提煉刑事算法開示的原則基礎上，構建算法開示制度的具體規則。

（一）刑事算法開示的構建原則

原則被視為一種體現著指令、標準和證成的指導性思想。^[110]考慮到前述算法開示構建的必要性、邏輯以及障礙等方面，應當從開示的邏輯、主體、方式以及階段等四個方面，著重遵循如下四個原則，以便為設計、解釋具體規則提供指導性方向和依據。

1. 權利保障優先原則

算法開示的構建邏輯存在權力行使邏輯與權利保障邏輯的分野，更進一步講，是國家權力與公民權利的分野。在制度設計之初，立法者需以現存問題為導向，錨定構建應然邏輯，避免制度異化與畸形。筆者認為，應當確立和遵循“權利保障優先原則”即權利保障邏輯優先於權力行使邏輯。原因如下：第一，權利保障邏輯為算法開示制度提供正當性。權利保障邏輯本身契合訴訟關照、平等武裝、訴訟參與等基本訴訟法理。同時，權利保障優先原則保與刑事訴訟價值具有一致性。隨著人權得到普遍認同，刑事訴訟法的功能也從工具法轉為保障法。^[111]“權利保障優先原則”可以保障法體系內部價值融貫，避免價值沖突。第二，權利保障邏輯更契合算法開示的制度目的。構建算法開示制度的目的，在於打破算法遮蔽式運作的實踐現狀。由於權力行使邏輯的追訴性，它依托於認罪認罰制度，使得庭審程序大幅簡化、證明對象轉變。它並沒有破除庭審中算法證據的概括式審查。權利保障邏輯通過保障被迫訴人的參與權、接受公正審判權，極大程度消解庭審中算法證據概括式審查的風險。因此，權利保障邏輯更契合算法開示的制度目的。第三，權利保障邏輯為權力行使提供兜底性保障。權力行使邏輯依托於認罪認罰制度，而認罪認罰過程中，存在協議破裂，從而導致程序反轉的可能。^[112]倘若認罪認罰協議破裂，訴訟程序需要從第四範式轉為第三範式，權力行使邏輯運行空間不復存在，亟待權利運行邏輯保障被迫訴人的基本訴訟權利。從運作空間而言，權利保障邏輯可以為權力行使邏輯兜底。

2. 析理開示為主原則

有學者將行政透明類型化為“魚缸型透明”（fishbowl transparency）和“析理型透明”（reasoned transparency）。前者以魚缸為隱喻，側重於公眾獲取政府掌握信息，主要解決的是政府正在做什麼。後者是指政府需要揭示它行動的原因，側重於說明為何如此行動的理由，主要解決的是政府為什麼做。這兩者並非截然區分，後者解釋程度依賴於前者。同時，實現兩種透明都是政府的義務。上述劃分思路，對算法開示具有重大的指導意義。

算法開示應當遵循“析理開示為主”原則。所謂“析理開示為主”原則是指司法機關或委托

[110] [德]卡爾·拉倫茨：《正確法：法倫理學基礎》，雷磊譯，法律出版社2022年版，第14-15頁。

[111] 參見魏曉娜：《從工具法到保障法：刑事訴訟法的功能轉型》，載《河北學刊》2024年第3期，第10頁。

[112] 馬明亮：《認罪認罰從寬制度中的協議破裂與程序反轉研究》，載《法學家》2020年第2期，第118頁。

其他人員解釋說明，算法數據來源、運行過程、以及如何得到分析報告等環節步驟。同時，析理開示為主，並不意味著排斥其他開示方法。當“析理開示”不能說服辯方或算法成為控辯雙方所爭論的焦點時，可以采用其他開示方案。確立“析理開示為主”原則的理由在於：第一，“析理開示為主”原則契合我國司法機關職能定位。上述提及，無論是“魚缸型開示”^[113]亦或是“析理型開示”，都是國家機關之義務。其中，“魚缸型開示”要求將算法源代碼、數據完全公開，但公開不同於理解，辯方需要聘請相關技術人員對算法源代碼進行解讀和理解。換言之，“魚缸型透明”將“理解成本”分配給當事人。“析理開示”則是將理解成本分配給司法機關，也契合我國司法機關的訴訟關照義務。第二，兼顧訴訟效率與權利保障。雖上述提及權利保障優先原則，但仍需考慮訴訟效率等因素。倘若過分強調算法源代碼、訓練數據公開，反而會出現數據、代碼傾倒，轉移訴訟爭點，降低訴訟效率等問題。因此，“析理開示為主”原則既可以解決上述問題，又兼顧保障被追訴的權利。第三，“析理開示為主”原則極大程度保護商業秘密和國家秘密。源代碼是商業秘密、國家秘密的核心，倘若動輒開示出來，無疑會對企業研發積極性、行業發展存在重大打擊。析理開示為主原則，要求司法工作人員解釋說明算法如何運作的，從而避免頻繁開示源代碼，損害商業秘密、國家秘密。但這並不意味著放棄了其他算法開示的方式。算法一旦成為訴訟焦點，訴訟效率需讓位於權利保障，便可以開示源代碼、訓練數據、先前測試數據等，確保被追訴人的權利。第四，“析理開示為主”極大程度規避算法複雜性導致的無效開示。析理開示並不探究源代碼的內容，極大程度避免了源代碼專業性、冗長性造成的無效開示。第五，“析理開示為主”原則要求司法工作人員對算法系統足夠了解。這一原則有利於增強司法工作人員責任心，督促司法工作人員主動揭開算法的神秘面紗。

3. 強制開示原則

強制開示原則要求司法機關將通過算法得到的證據材料全部入卷並且開示，否則證據材料不具備證據資格。強制開示原則具有三項子原則構成：第一，全部入卷原則。實踐中，偵查機關通過預測性算法得到的部分證據材料並不放入卷宗。辯護人無法通過閱卷獲取相關信息，大大減弱了被告人辯護效果。因此，需要確立全部入卷原則，即司法機關將通過算法得到的證據材料全部放入案卷之中。第二，全面開示原則。針對案卷中通過算法得到的證據材料，司法機關應當對其算法全面開示，開示方式可以參照析理開示為主原則。第三，制裁原則。徒法不足以行，若缺乏必要的制裁性法律後果，會導致強制開示原則虛置，妨礙強制性的實現。同時，設置喪失證據資格的制裁原則，可以倒逼司法機關工作人員開示算法。

有學者提出“必要性原則”，即只有控辯雙方對算法科學性問題產生激烈爭議，一方當事人向法院提出公開算法的請求，並且法院也認為有必要公開算法時，才需要公開。^[114]這種觀點，有待商榷。一方面，“必要性原則”可行性不強。“必要性原則”無法指導法律規則明確啟動開示條件、相應的證明標準等要件。最終，算法開示啟動與否，淪為法官自由裁量之程序。這可能會造成弱化程序剛性，虛置程序等後果。另一方面，“必要性原則”將主要證明負擔轉移給辯方，拉大控辯力量落差。由於控辯雙方算力不對等，實踐中，控方提供算法證據的情況較多。相較於控方，絕大部分案件，辯方更希望開示算法。此時，需要辯方負擔證明“有必要”的責任，並且需要達到說服法官的程度。此種理解加大了算法開示難度，增加給辯方的負擔，進一步加劇了控辯雙方的不對等。

[113] 與魚缸型透明的概念相對應。

[114] 洪濤：《大數據證據真實性審查規則的建構》，載《蘇州大學學報（法學版）》2024年第1期，第78頁。

4. 全程開示原則

全程開示原則要求司從算法設計到審判階段，立法者都需要設計相應程序，開示算法。確立全程開示原則，可以從三個維度展開論證：宏觀維度，全程開示原則可以吸納多元主體參與。從算法設計環節，設計者對算法的設計思路、相關參數予以說明、解釋，並借由傳統媒體、網絡平台傳播，形成輿論聚合，吸納除刑訴主體外更多數人的意見。盡可能消除算法偏見、統一算法內容。中觀維度，全程開示原則可以全流程保護被告人訴訟權益。相較於算法全流程的遮蔽式運用，奉行全程開示原則可以進一步保障被告人知情權、對質權等訴訟權益。微觀維度，全程開示原則可以修正閱卷權的局限性。根據我國刑訴法，審查起訴階段，辯護律師才能進行閱卷。全程開示原則補足了偵查階段，甚至初查階段的信息獲知空白，更好的保障被告人權益。

（二）刑事算法開示的具體制度構建

算法開示制度應當遵循上述四項原則，同時，結合不同流程環節，不同訴訟主體、開示內容、法律後果的差異，構建刑事算法開示的具體制度。

1. 算法開發環節的開示

算法開發環節的開示主體應當是偵查機關；開示內容包括設計算法的目的、考慮的要素、要素之間的關係、識別的誤差率等；若不予開示，則通過該算法分析得到的證據材料，不具有合法性。算法開發環節明顯早於刑事訴訟發生，但開發環節所設計的算法產品，將對刑事訴訟產生極大影響，因此，有必要在開發環節將設計理念、參數以及設計標準予以開示，避免後續應用算法時，存在偏見執法與偵查。同時，設計環節開示，可以有效吸納多元主體參與，極大程度破除算法設計中“弱二元主體”模式。同時，按照強制性開示原則，還應當設計相應的法律後果，即取得的證據材料不具有合法性，應當予以排除。此種合法性審查規則，也可以為後續法庭審查、律師辯護提供法律依據。

2. 立案階段的開示

立案階段開示主體是偵查機關；開示內容應當包括：偵查機關利用算法查明犯罪事實或犯罪嫌疑人得到的全部證據材料、算法的功能與適用場景、算法識別的誤差率等；開示方式是析理開示；若不予以開示，通過算法分析得到的證據材料不具有證據資格。算法、大數據技術偵查可能使立案程序虛置化，加之公安機關行政權、偵查權混淆。算法適用時間節點前移至立案前，“早在刑事訴訟正式啟動之前，證據就以情報這一形式被提前固定，一旦偵查機關啟動刑事程序即可瞬時完成取證行為。”^[115]因此，需要在立案階段結束後，偵查階段開示前，進行算法開示，這也符合實踐中算法應用特點。根據我國《刑訴法》第109條，刑事立案條件是“發現犯罪事實或犯罪嫌疑人”。因此，對於立案階段存在犯罪嫌疑人的案件，應當對其進行算法開示；對於僅存在犯罪事實，不確定犯罪嫌疑人的案件，立案階段不需要進行算法開示。另外，立案階段直接決定公民是否可能進入刑事訴訟程序之中，因此需要較為嚴格的程序控制，不能僅因算法或機器提供分析意見，便將公民視作潛在犯罪嫌疑人。如果犯罪嫌疑人有所異議，應當記錄在案。

3. 偵查階段的開示

偵查階段開示主體是偵查機關；開示內容應當包括：偵查機關利用算法得到的全部證據材料、算法功能與適用場景、鑒定所用的算法程序名稱、功能介紹等；開示方式是析理開示；若不予以開示，通過算法分析得到的證據材料不具有證據資格。偵查階段進行算法開示具有規範依據。根據《刑訴法》第162條以及《公安規定》第289條規定，案件偵查終結、移送檢察機關同時需要將案件移送情況告知犯罪嫌疑人與辯護律師。在此階段，可以進行算法開示。同時，根據《公安規定》

[115] 李訓虎：《偵查情報化之批判》，載《法學》2024年第7期，第132頁。

第 287 條，偵查機關移送案卷時，僅移送訴訟卷，而不移送偵查卷。偵查機關應當遵循強制開示原則，將算法所得全部證據材料，都放入訴訟卷，供檢察機關和辯護律師查閱。另外，根據《刑訴法》第 148 條，偵查機關應當將用作證據的鑒定意見告知犯罪嫌疑人、被害人。因此，鑒定環節中所使用的算法也應當進行開示。倘若由於鑒定應用算法專業性過強或偵查人員並不熟悉相關算法，偵查機關還可以委托鑒定人員進行開示。如果犯罪嫌疑人有所異議，應當記錄在案。另外，倘若辯方也有通過算法得出

4. 審查起訴階段的開示

在審查起訴階段，檢察機關的自行補充偵查和與犯罪嫌疑人協商簽署認罪認罰協議，需要進行算法開示。針對檢察機關自行補充偵查，筆者認為可以參照既有制度安排，將算法開示活動置於庭前會議之中，內容與法律後果可以參照偵查階段的開示。針對協商簽署認罪認罰協議，由於其關乎被追訴人是否放棄公正審判權以及需要保障被追訴人的自願性，需要在審查起訴階段進行算法開示。

審查起訴階段開示主體是檢察機關；開示內容應當包括：利用算法得到全部證據材料外；算法功能與適用場景、算法涉及被告人全部個人數據等；開示方式是析理開示；若未經開示，犯罪嫌疑人簽署的認罪認罰協議無效，案件後續不適用認罪認罰程序。

5. 審判階段的開示

審判階段的開示可以分成：庭前會議的開示以及庭審中的開示。

庭前會議階段的開示。根據我國《人民法院辦理刑事案件庭前會議規程》第 2 條，庭前會議可以由人民法院組織控辯雙方展示證據，歸納爭議焦點。此處，雙方可以就算法運行、數據來源等有關問題，進行充分探討。開示主體是檢察機關，開示內容包括：案卷中由算法得到的證據材料，算法功能與適用場景，數據來源等；開示方式是析理開示；倘若辯方對此並無異議，應當參照《人民法院辦理刑事案件庭前會議規程》第 25 條，達成一致意見後，庭審中反悔，除由正當理由外，法庭一般不再處理；倘若辯方對算法有異議，應當予以提出，經法官審查後，認為確有必要，可以進行庭外源代碼、數據公開。

庭審中的開示。由於算法證據的概括式審查，庭審中，算法仍遮蔽式地運作。因此，在裁判文書中，法官需要重點論證，審查算法證據材料的關聯性與真實性的過程和思路。倘若，法官仍以概括式審查遮掩算法運行，則可以作為二審上訴理由或者再審理由，提起上訴或再審。

Abstract: In criminal proceedings, algorithms are often used in a concealed manner, manifested by the dual black boxes of algorithms in the investigation process, the implicit recognition of algorithms in the forensic identification system, and the cursory examination of algorithmic evidence during trials. This concealed use of algorithms may lead to risks such as the encapsulation and fragmentation of algorithm content, the absence of technical due process, and increased dependence among judicial personnel. It is necessary to establish a formal disclosure system for criminal algorithms to facilitate the formation of unified and improved standards, establish the foundation for criminal technical due process, and promote algorithm accountability. The establishment of the algorithm disclosure system needs to follow two logics: the logic of exercising power based on official duties and serving litigation efficiency, and the logic of safeguarding rights by ensuring litigation participation and the right to a fair trial. In the process of constructing the algorithm disclosure system, it is necessary to properly handle issues such as conflicts between algorithm disclosure and commercial or state secrets, ineffective disclosure due to algorithm complexity, and obstacles posed by traditional disclosure systems to algorithm disclosure. To establish the criminal algorithm disclosure system, it is essential to clarify the principles of prioritizing rights protection, focusing on analytical disclosure, mandatory disclosure, and full-process disclosure. These principles should be implemented throughout the stages of algorithm development, case filing, investigation, prosecution review, and trial, with specific procedures established accordingly.

Key words: Algorithm Disclosure; Algorithm Transparency; Analytical Disclosure; Technical Due Process

(責任編輯：張雨虹)