

刑事司法人工智能的包容性规制^{*}

李 训 虎

摘 要：我国刑事司法人工智能走在世界前列，但暴露出数据垄断、算法黑箱以及应用场景设置随意等问题。这些问题及其背后所面临的合法性挑战、正当性隐忧和伦理性风险的科学解决与化解，要求超越传统思维，创新监管模式，引入技术赋权理念和技术正当程序，对刑事司法人工智能进行包容性规制，以实现发展与规制的协调。刑事司法保障人权与人工智能以人为本的深度融合形成技术赋权理念，可以为刑事司法人工智能的研发、应用提供理念层面的宏观指引；在传统正当程序基础上发展而来的技术正当程序，可以对刑事司法人工智能应用进行过程规制，以促进司法数据的公开，提升算法的透明度，并推动刑事司法人工智能问责机制的建立。数据驱动下的司法人工智能应用对司法结构和司法治理模式形成冲击，并对司法运行产生重塑效应，逐渐形成人机协同的司法治理新模式。未来应当秉持以人为本、技术赋权理念，重塑人机关系新格局下的价值取向和行为准则，遵循法律公平正义的价值追求，构建人机协同司法治理新格局。

关键词：刑事司法 人工智能 技术赋权 技术正当程序 规制

作者李训虎，中国政法大学证据科学研究院教授（北京 100088）。

引 言

随着 2016 年 AlphaGo 的横空出世，人工智能迅速升温。在各国人工智能战略和资本市场的强势推动下，以深度学习为主要特征的新一轮人工智能浪潮席卷全球，成功赋能传统产业并对其产生深刻影响。随着人工智能赋能司法领域，以智能辅助办案系统为代表的刑事司法人工智能开始崛起，对传统司法业务形成巨大冲击，引发司法权力运行机制、司法行为运作方式的深刻变革，^① 并对既有的司法伦理、正义理念构成挑战，^② 而

* 本文为中国政法大学钱端升杰出学者支持计划资助项目阶段性成果。

① 参见季卫东：《人工智能时代的司法权之变》，《东方法学》2018 年第 1 期。

② 参见马长山：《司法人工智能的重塑效应及其限度》，《法学研究》2020 年第 4 期；伊森·凯什、奥娜·拉比诺维奇·艾尼：《数字正义：当纠纷解决遇见互联网科技》，赵蕾等译，北京：法律出版社，2019 年。

现行刑事诉讼法却缺乏有效规制手段。

对刑事司法人工智能的规制，各国都处于起步阶段，都在探索适合本国国情的规制路径，形成发展优位的中国模式、^① 严格监管的欧盟模式^②和渐进规制的美国模式，^③ 呈现刑事司法人工智能规制的国际竞争格局。^④ 虽然我国刑事司法人工智能应用走在世界前列，但规制却处于缺位状态，长此以往会陷入“想控制却难以控制”的科林格里奇困境（Collingridge's dilemma）。^⑤ 未来应按照习近平总书记提出的“推动大数据、人工智能等科技创新成果同司法工作深度融合”“让人民群众在每一个司法案件中感受到公平正义”的法治思想，改进发展优位的规制模式，构建专门针对刑事司法人工智能的规制原则、规制框架和规制规范，防范、化解可能存在的风险和挑战，通过激励发展与合理监管相协调实现包容性规制。

本文将通过考察我国刑事司法人工智能的应用现状，展示发展成果，揭示存在的实际问题，分析面临的合法性挑战、存在的正当性隐忧和潜存的伦理性风险，阐释刑事司法人工智能的规制原理，提出现实可行的规制方案。结语部分，进一步审

- ① 最高人民法院、最高人民检察院相继发布系列文件，推动人工智能与刑事司法的融合，将创新、发展刑事司法人工智能作为主旨，形成发展优位的刑事司法人工智能推进思路，刑事司法人工智能的规制基本处于缺位状态。
- ② 欧盟通过宏观的人工智能战略与具体的刑事司法人工智能规制措施建立起严格的刑事司法人工智能监管体系：《通用数据保护条例》通过对于数据的规制建立源头监管；《算法责任与透明治理框架》《人工智能白皮书》强调对于算法的规制；欧洲司法效率委员会则通过系列文件划定刑事司法人工智能的发展空间与具体实现路径。参见 European Commission for the Efficiency of Justice, “European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment,” Council of Europe, 2018.
- ③ 美国人工智能政策以促进人工智能技术和产业发展、降低创新门槛和成本为优先考虑，并在发展过程中采取渐进规制模式来应对新出现的问题。受此影响，在刑事司法人工智能领域，美国法院和州立法机构分别通过判例和法案的形式对具体刑事司法人工智能应用进行规制。参见曹建峰：《人工智能治理：从科技中心主义到科技人文协作》，《上海师范大学学报》2020年第5期。
- ④ 《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》提出“积极参与数字领域国际规则 and 标准制定”；英国上议院坦承其无力在人工智能领域与中美两国全面竞争，主张将重心聚焦于人工智能发展与应用的伦理领域，进而主导人工智能伦理领域的规则制定。参见 Select Committee on Artificial Intelligence, “AI in the UK: Ready, Willing and Able?” Authority of the House of Lords, April 16, 2018, p. 120.
- ⑤ 科林格里奇困境展示了人类对技术进行规制面临的信息和能力困境。技术在被广泛应用前，由于难以认知其有害社会后果，此时可以对其控制却不知如何控制；随着技术发展，当不可欲的后果发生时，技术已经根深蒂固，成为整个社会结构的一部分，已经很难对其进行控制。参见 David Collingridge, *The Social Control of Technology*, London: Frances Printer, 1980, p. 19.

视新一轮科技革命背景下人机关系的变革对既有司法伦理、正义理念的冲击，反思人机协同的司法治理新模式。

一、中国刑事司法人工智能应用考察

对于快速发展的刑事司法人工智能应用，我国学术界缺乏深入研究，其对于刑事司法人工智能的科学总结与学理思考，相较于司法实务界与科技界倾注的努力，不相匹配，亦未提出具有针对性的规制方案。全景展示我国刑事司法人工智能应用的发展现状，揭示存在的问题，是对刑事司法人工智能进行科学规制的前提和基础。

（一）走在世界的前列

对于人工智能在刑事司法领域的应用，欧美国家司法实务机关普遍持相对保守态度。美国将刑事司法人工智能的应用场景限定于预测性警务、人脸识别以及审前程序、量刑程序中的风险评估。^①英国在审前程序中的预测性警务、风险评估、移动设备信息提取、数字取证等领域进行刑事司法人工智能应用试点。^②德国警方除曾试用人脸识别、在个别地区使用预测性警务外，基本没有引入刑事司法人工智能应用；^③法国禁止将基于人工智能技术的法律服务软件应用于刑事案件，并且通过立法禁止基于法官和书记官处成员的身份进行的人工智能裁判指引。^④总体而言，欧洲主要国家只是在审前程序有个别刑事司法人工智能应用，审判阶段几乎不存在刑事司法人工智能应用。^⑤

① 参见 Arthur Rizr and Caleb Watney, "Artificial Intelligence Can Make Our Jail System More Efficient, Equitable and Just," *Texas Review of Law & Politics*, vol.23, no.1, 2019, p.183.

② 参见 The Law Society Commission on the Use of Algorithms in the Justice System et al., "Algorithms in the Criminal Justice System," *The Law Society of England and Wales*, June, 2019.

③ 参见 Lukas Staffler and Oliver Jany, "Künstliche Intelligenz und Strafrechtspflege-eine Orientierung," *Zeitschrift für Internationale Strafrechtsdogmatik*, 2020, S.167-176; Veronica R. S. Hoch, "Big Data und Predictive Analytics im Gerichtsprozess. Chancen und Grenzen der Urteilsprognose," *Zeitschrift für IT-Recht und Recht der Digitalisierung*, 2020, S.296-300.

④ 参见 Gary E. Marchant, "Emerging Technologies and the Courts," *Court Review*, vol. 55, no. 4, 2019, p.152.

⑤ 参见 European Commission for the Efficiency of Justice, "European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment."

与欧美国家不同的是,我国各级各地司法实务机关积极研发、应用刑事司法人工智能,涵盖刑事审前程序、定罪程序及量刑程序,几乎实现对刑事诉讼流程的全覆盖。^①我国《新一代人工智能发展规划》明确提出,“建设集审判、人员、数据应用、司法公开和动态监控于一体的智慧法庭数据平台,促进人工智能在证据收集、案例分析、法律文件阅读与分析中的应用,实现法院审判体系和审判能力智能化。”因应国家层面推出的人工智能顶层设计,各级各地司法实务机关积极推动人工智能落地生根,最高人民法院提出充分运用司法大数据,依托人工智能,以人工智能引领“智慧法院”建设,着力推进审判、执行、服务智能化;最高人民检察院力图打造数据化、科学化、智能化的“智慧检务”,提出构建“全业务智慧办案、全要素智慧管理、全方位智慧服务、全领域智慧支撑的智慧检务”总体架构;地方司法实务机关积极将“大数据、人工智能与司法体制改革结合起来”,短期内诸多具体刑事司法人工智能应用相继投入应用,我国成为全球刑事司法人工智能应用领域的先行者。

在我国的刑事司法人工智能建设实践中,逐渐发展出四种类型的刑事司法人工智能应用:^②(1)智能辅助决策应用系统。司法实务机关将人工智能应用于刑事审判中的证据审查、事实认定、法律判断以及审前程序中的逮捕社会危险性评估等领域,以帮助决策者作出判断。上海市高级人民法院的“刑事案件智能辅助办案系统”为其中的典型代表。(2)智能辅助支持系统。司法机关引入立案风险自动拦截、敏感案件自动标识预警、庭审语音识别、辅助录入、辅助阅卷、文书智能纠错、类案强制检索等以智能为核心的关键技术,以提升检察、审判质效。(3)案件管理应用系统。通过引入刑事司法人工智能提高诉讼效率是改革者的重要诉求,例如,检察机关将“全要素智慧管理”作为“智慧检务4.0”的关键一环,力图“通过探索建立基于管理决策大数据的新一代人工智能检察管理模式,统筹优化检察机关内部‘人、事、财、物、策’各项管理要素,全面提升检察机关现代化水平。”^③在江苏省检察系统获得广泛应用的“案管机器人”为其中的典型代表。(4)诉讼服务应用系统。该类人工智能以方便当事人诉讼、为当事人提供法律咨询为主要特色,旨在为当事人提供方便快捷的诉讼服务。

司法实践中,司法实务机关推出的刑事司法人工智能应用往往兼具辅助决策、辅助支持、案件管理和诉讼服务等多种功能,目前已经落地的刑事司法人工智能应

① 参见崔亚东:《人工智能与司法现代化》,上海:上海人民出版社,2019年,第109页。

② 由于公安机关刑事司法人工智能的信息披露相对匮乏,本文主要以法院、检察院用于辅助决策、办案的刑事司法人工智能应用为分析对象。

③ 田野:《智慧检务开启司法文明新纪元》,《清风苑》2018年第3期。

用已经达到数百个之多。^①

（二）先行先试的探索

自2017年全国司法体制改革推进会正式提出司法改革要“主动拥抱大数据、人工智能时代”，“推广应用智能辅助系统，探索构建人力和科技深度融合的司法运行新机制”以来，刑事司法人工智能建设已经取得初步成效。但遗憾的是，针对刑事司法人工智能的既有评估多为实务部门的自我评价，对智慧审判、法院信息化建设进行的第三方评估则缺乏对于刑事司法人工智能的专门关注。本文力图通过对上海、北京、贵州、山西等地刑事案件智能辅助办案系统运行情况的实证调研，管窥当下的刑事司法人工智能运行生态。

上海市高级人民法院的刑事案件辅助办案系统围绕证据的有效收集固定、保存和审查判断以及规范公检法办案流程来展开。^②根据实务部门的自我评价，该系统的实施效果主要体现在：统一网络平台，实现互联互通；统一证据标准，实现标准化、规范化、数据化；实现校验、把关、提示，有效防止冤假错案；多项功能集成应用，节约资源成本，提升办案质效等。^③接受访谈的检察官、法官更多地将上述评价作为该系统未来努力的方向和目标，普遍认为，该系统的主要作用在于促进办案信息化并提高诉讼效率，但无法进行证据校验和审查判断，目前的技术无法支撑其参与复杂的事实认定工作，且复杂案件中侦查阶段难以完成繁杂的数据标注和扫描工作严重制约了系统的适用。该系统在山西太原等地法院的实施情况为上述判断提供了有力注脚。^④

北京法院系统力图构建服务统一裁判尺度、提供统一办案指引、智能支持办案全流程的“睿法官”系统，力图实现自动生成“案情画像”，生成审理环节提纲、笔录模板以及结案环节的裁判文书。^⑤然而，由于数据孤岛、专业壁垒、投入壁垒和

① 以智慧检务为例，全国各级各地检察机关已经推出两百余个智慧检务应用。参见赵志刚、金鸿浩：《智慧检务概论——检察机关法律监督的科技智慧》，北京：中国检察出版社，2018年，“前言”，第2页。

② 参见黄祥青：《“206工程”的构建要点与主要功能》，《中国检察官》2018年第15期。

③ 参见崔亚东：《人工智能与司法现代化》，第184—192页。

④ 受访法官表示，该系统设计过于理想化，缺乏技术支撑，设计功能难以实现。该系统在应用中的主要作用在于方便阅卷、方便撰写判决书时摘录信息，但并没有体现出所谓的智能辅助办案功能。另外，阅卷及信息提取功能的发挥还受制于侦查机关是否将侦查卷扫描录入系统，在侦查卷数量众多的案件中，侦查机关没有精力从事如此繁重的工作，导致该系统处于闲置状态。

⑤ 参见曹雅静：《搭建现代化信息桥梁 推进“智慧法院”建设》，《人民法院报》2016年12月15日，第4版。

融合壁垒以及系统学习能力等方面存在欠缺,^①“睿法官”系统并未得到真正应用。对贵州某法院法官的访谈则显示,前期备受关注的贵州刑事案件智能辅助办案系统基本没有在刑事审判中得到应用,甚至有受访法官对此持完全的否定态度,认为该系统仅仅处于起步阶段,根本无法达到预期效果。

作为刑事诉讼重要参与者的辩护律师普遍反映,他们并不知晓刑事案件智能辅助办案系统在当地实践中的应用,普遍担心,该系统在公检法三机关之间实现互联互通,有助于强化配合、弱化制约,加剧本已失衡的控辩关系,降低当事人权利的保障水平。

调研还显示,由于科技公司、司法实务机关对刑事案件智能辅助办案系统为代表的刑事司法人工智能可能的风险缺乏足够的认知,导致研发、应用环节并未建立相应的责任追究机制。

尽管不能因为刑事案件智能辅助办案系统存在诸多问题,就对整体的刑事司法人工智能予以简单的批判或者否定,但作为实践中获得广泛关注并得到政法高层大力推广的刑事案件智能辅助办案系统,否定性评价的存在促使我们反思,未来的刑事司法人工智能究竟如何推进。

(三) 发展的基础问题

刑事案件智能辅助办案系统的应用、推广未达预期,是多元、复数因素共同作用的结果,但其中发挥决定作用的是刑事司法人工智能应用本身是否契合刑事司法规律,所依托技术是否具有可行性。由此,需要从决定刑事司法人工智能应用的数据、算法以及应用场景三个核心要素展开分析。

1. 数据垄断: 制约刑事司法人工智能的基础难题

刑事司法人工智能的快速发展得益于大量可以用于训练分析的大数据资源、设计精巧的算法以及能够支持超大规模计算的算力的支持。海量的司法案件为刑事司法人工智能应用奠定了坚实的数据前提和基础。但遗憾的是,尽管我国司法信息公开、司法大数据平台建设有目共睹,但刑事司法大数据依然存在诸多制约刑事司法人工智能发展的缺陷。

首先,数据孤岛、数据壁垒现象严重。公检法机关都掌握海量刑事司法数据,其他政府机关、社会机构亦掌握大量相关数据。但刑事司法机关之间、刑事司法机关与相关机关之间并没有真正实现数据的互联互通,当数据孤岛、数据壁垒现象难以解决时,研发人员只能依据本部门数据或者最终环节的刑事裁判文书进行数据分

^① 参见余贵清等:《借力大数据 智慧助法官——北京法院睿法官系统》,李林、田禾主编:《法治蓝皮书:中国法院信息化发展报告 No.2 (2018)》,北京:社会科学文献出版社,2018年,第349页。

析,这无疑会在一定程度上限制刑事司法人工智能应用的质量。另外,各级各地司法机关在推进刑事司法人工智能方面基本处于各自为政的状态,尽管这些自发改革有促进技术竞争的积极作用,但竞争的存在也使得相关司法机关之间人为制造壁垒,数据难以有序流动,难以实现共享。

其次,数据缺失、数据瑕疵问题突出。中国裁判文书网已经成为全球最大的司法大数据平台,但大量裁判文书所载信息,特别是分析、论证过程相对简略,对于研发高质量的刑事司法人工智能远远不够。并且,真正记录法官分析过程和实质裁判要素的合议庭笔录、审理报告等内部文书,尚无法成为研发人工智能的底层数据,使得研发人员只能借助最终裁判文书中记载的有限信息展开分析,这无疑会严重影响机器学习的质量。同时,囿于选择性发布、部分发布的存在,民众无法获取全部信息;即便是裁判文书公开较为积极的省份,亦存在比较明显的数据漏洞问题,^①这无疑会影响刑事司法人工智能的研发质量。另外,刑事司法数据存在的瑕疵问题同样值得关注。正如学者所言,“在由小数据时代向大数据时代转变的过程中,我们对信息的一些局限性必须给予高度的重视。数据的质量可能会很差;可能是不客观的;可能存在分析错误或者具有误导性;更糟糕的是,数据可能根本达不到量化它的目的。”^②

最后,数据垄断、数字鸿沟难解。虽然存在数据孤岛、数据壁垒问题,存在如何通过人工处理将纷繁复杂的信息转化为机器能够识别的语料问题,以及如何让机器准确读取、抓取有用信息等技术问题,但这些问题依靠技术进步可以在一定程度上得以缓解。当下更为根本、更具决定意义的问题是,刑事司法机关的数据垄断和数据独占问题。由于我国的刑事司法大数据基本处于刑事司法机关的完全控制之下,且未建立完善的信息公开和发布机制,使得民众无法获取对等有效的信息,从而形成难以逾越的“数字鸿沟”,使得刑事司法机关在大数据加持下力量得到持续强化,进而使控辩双方之间、民众与审判机关之间产生信息不对称问题。当刑事司法数据处于封闭、垄断状态,民众缺乏有效获取途径时,这不仅会侵害民众的知情权,也会侵害民众对于司法的认同度。而且,数据垄断的存在会在相当程度上阻断科技公司独立研发高质量刑事司法人工智能的可能性,从而使得刑事司法机关通过数据垄断建立起其他社会机构难以逾越的行业壁垒。

2. 算法黑箱:困扰刑事司法人工智能的基本难题

因诸多有意或无意的偏见及信息不充分等因素可能影响决策的公正性,人类发

① 参见中国社会科学院法学研究所法治指数创新工程项目组:《中国法院“司法大数据”应用第三方评估报告(2019)》,陈甦、田禾主编:《法治蓝皮书:中国法院信息化发展报告 No.4(2020)》,北京:社会科学文献出版社,2020年,第75—77页。

② 维克托·迈尔-舍恩伯格、肯尼思·库克耶:《大数据时代》,盛杨燕、周涛译,杭州:浙江人民出版社,2013年,第210页。

展出一种利用数学方法将人类社会事务量化、客观化的思潮，利用算法、模型、机器学习等数学方法，试图重塑一个更加客观的现实世界。^①但算法应用的深入推广使得人类已经处于一个被“算法吞噬的世界”，不得不面临由此带来的算法歧视、算法偏见以及算法黑箱问题。^②作为编程人员主观选择、判断结果、人脑外延的算法，会不可避免地融入编写人员的价值偏好，使得算法偏见、算法歧视无法避免，成为算法的天然属性。算法黑箱的形成则带有明显的人为属性，取决于算法设计者、实际控制者等主体是否主动公开算法。在我国，算法黑箱问题并没有因为大量刑事司法人工智能应用由司法机关主导研发而得到解决。

在刑事司法人工智能应用过程中，算法黑箱引发的不透明、不准确、不公平、难以审查等问题，^③备受质疑。具体言之，由于算法黑箱的存在，外界无从知晓刑事司法人工智能应用在具体场景中到底发挥多大作用，这会加剧外界对刑事司法人工智能的不信任。即便刑事司法人工智能只是发挥辅助作用，算法的参与也使得原有的决策模式发生变化。如若因算法存在偏差导致错误出现，就会产生问责与监管的难题。再者，当事人对影响诉讼结果的因素有权利提出救济诉求，但算法的参与会产生当事人难以质疑的隐性因素，诉讼的公平和正义无法获得保障。因算法黑箱导致的诸如此类的问题，如果得不到妥善解决，不仅会影响刑事司法的公信力，而且会影响具体应用者对于刑事司法人工智能的认同度。美国“量刑算法合宪性第一案”^④引发的广泛争议，凸显了算法黑箱问题已经成为制约刑事司法人工智能的基本难题。

3. 场景自设：实践刑事司法人工智能的激进尝试

我国地方司法机关推出的刑事司法人工智能具有广泛的应用场景，涵盖逮捕条件审查、文书自动生成、类案推送、裁判偏离度预警、辅助量刑等领域，并且特别重视人工智能在证据标准、证据规则指引、证据链和全案证据审查判断等事实认定领域中的应用，基本实现对刑事诉讼流程的全覆盖。这给人应用场景设置随意、缺乏规制的印象。

从全球刑事司法人工智能发展来看，运用刑事司法人工智能从事庭审录入、法条查询等重复性工作，从事诉讼服务、流程管理等事务性工作，已经被普遍接受。对于人工智能从事风险评估尽管存在不同认识，但总体上获得认同。对于刑事司法

① 参见腾讯研究院等：《人工智能：国家人工智能战略行动抓手》，北京：中国人民大学出版社，2017年，第241页。

② 参见 B. Bodo et al., “Tackling the Algorithmic Control Crisis—the Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents,” *Yale Journal of Law and Technology*, vol.19, no.1, 2017, pp.136-138.

③ 参见李开复、王咏刚：《人工智能》，北京：文化发展出版社，2017年，第240页。

④ 参见 *State v. Loomis*, 881 N.W.2d, 761.

人工智能应否进入事实认定领域,进行证据校验、证据审查判断,则存在普遍一致的反对声音。我国司法实务机关将事实认定领域作为刑事司法人工智能落地的重心,值得认真研判。

事实认定过程是充满价值判断和价值选择的过程,要求事实认定者秉持经验、良心和理性作出裁决。而人工智能的引入,则存在一个巨大疑问:预先确定的决策程序是否可以真正纳入人类情感因素,是否可以在刑事案件中建立对事实敏感因素的平衡机制。^①将人工智能引入事实认定领域,会使得复杂的事实认定过程简单化、机械化,使得本应被纳入综合评判的因素(权利、利益、损失、危害程度等)被弱化。特别是在弱人工智能时代,将人工智能应用于事实认定,还存在诸多难以克服的障碍。数据识别率不稳定、案件样本太少、数据分析不准等现实问题的存在,^②以及人工智能在证据数据化、数据整合和结论输出等环节面临的问题,都使其难以胜任事实认定工作。^③既有从事事实认定的刑事案件智能辅助办案系统实施效果,与制度设计者预期目标相去甚远,彰显了问题的严重性。这促使我们反思,刑事司法引入人工智能的限度,以及未来刑事司法人工智能应用场景设置的边界。

我国各级司法机关在刑事司法人工智能领域积极探索,走在世界的前列,但迅速发展的刑事司法人工智能逐渐暴露出数据垄断、算法黑箱以及场景设置随意等问题。实证调研亦充分显示,刑事案件智能辅助办案系统作为刑事司法人工智能的集大成者在推广、试点过程中运行并不尽如人意。正视改革中的问题,探寻背后的深层制约因素,成为建构刑事司法人工智能规制方案的基础。

二、中国刑事司法人工智能应用的法理难题

刑事司法人工智能应用过程中存在的问题表明,其只是现代科技融入刑事司法的一种技术进步力量,应当以促进司法正义实现的工具或辅助手段形式存在,^④其背后的价值基础是司法正义。为此,刑事司法人工智能应用不仅需要具备法律层面的合法性,更需要法理层面的正当性,以克服潜在的伦理风险。由此,从刑事诉讼

① 参见 Richard M. Re and Alicia Solow-Niederman, "Developing Artificially Intelligent Justice," *Stanford Technology Law Review*, vol.22, no.2, 2019, p.253.

② 参见杨敏:《上海刑事案件智能辅助办案系统调研报告》,李林、田禾主编:《法治蓝皮书:中国法院信息化发展报告 No.2 (2018)》,第198页。

③ 参见栗峥:《人工智能与事实认定》,《法学研究》2020年第1期。

④ 2017年一篇“南京市中级人民法院将引入机器人判案”的媒体报道引发“机器人法官之争”,尽管南京中院及时回应,澄清“人工智能代替法官判案”的误解,但依然引发广泛热议。

法律规范层面对刑事司法人工智能应用进行合法性检验，从刑事诉讼法理层面进行正当性评析，从刑事司法正义结构层面进行整体性评估，方可发展出契合中国刑事司法正义目标的人工智能应用。

（一）面临的合法性挑战

首先，事实认定领域的刑事司法人工智能应用会侵犯事实认定者的法定权力。根据刑事诉讼法相关条文，特别是第3条、第183条、第184条、第185条、第198条、第200条及第203条的规定，审判权由人民法院行使，事实认定者只能为法官和人民陪审员。在既往刑事司法实践中，庭审虚化现象严重，以审判为中心的刑事诉讼制度改革着力推进庭审实质化，力图真正实现“让审理者裁判、由裁判者负责”，法官和人民陪审员作为事实认定者的身份更加凸显。在这一背景下，应用于事实认定领域，特别是具备“证据标准指引、单一证据校验、证据链和全案证据审查判断”功能的刑事司法人工智能应用，是否会侵犯法律规定的事实认定者权力，是否会导致庭审空心化，阻碍庭审实质化改革目标顺利实现，值得深入反思。

有论者认为，刑事司法人工智能只是辅助工具，有助于准确认定事实、能够促进法律统一适用。现实的问题在于，通过证据数据化、法律规则代码化对审判人员进行裁判指引，无异于通过层层嵌套的技术实现对司法人员裁判权力的削弱，一定程度上会产生决策让渡现象。并且，在案件数量剧增、审理时间严格受限的压力之下，加之刑事司法人工智能具备判断效率上的比较优势，运用人工智能的法官难免会形成依赖惯性甚至惰性，长此以往势必会形成过度依赖人工智能的现象。如此一来，司法实践中的事实认定者将不再只有法律规定的法官，而演变为数据分析师、人工智能工程师、软件工程师与法官的集合体，裁判权也将异化为计算能力和裁判能力的融合运用，这无疑会侵蚀现行法律授予法官的事实认定权，亦会侵犯法律赋予人民法院独享的审判权。

其次，刑事司法人工智能应用的引入可能会侵犯被追诉人的诸多法定权利。预测性警务等犯罪预测工具的引入，不仅会使得追诉活动开始于正式的刑事立案之前，而且人工智能加持下的海量信息汇总、分析产生的犯罪画像，会加剧侦查人员形成有罪推定的判断，并使得这一判断通过互联互通的刑事案件智能辅助办案平台传递至法官，导致法官形成未审先判的预断，庭审实质化流于形式，构成对被追诉人无罪推定权利的实质侵犯。根据刑事诉讼法规定，当事人有知悉诉讼进程、参与诉讼过程、参与举证、质证以及法庭辩论的一系列诉讼权利，但刑事案件智能辅助办案系统的引入，使得原先当事人具有的针对审判人员或者对方当事人的诉讼权利都难以适用于人工智能裁决系统，导致当事人诉讼权利的实质受损。

再次，风险评估类刑事案件智能辅助办案系统可能会加剧已有的偏见。在刑事司法领域引入人工智能对被追诉人的社会危险性、再次犯罪的可能性进行风险

评估，以准确作出是否逮捕、是否继续羁押的决定时，看似只是所运用技术的迭代升级，但域外实践已充分证明，风险评估工具的运用会造成对特定群体的歧视，导致其被羁押、被从重量刑的概率显著增加。^① 进而言之，刑事司法人工智能应用存在加剧或者强化已有偏见、歧视的副作用。刑事司法人工智能的研发需要依托司法大数据，但当底层数据存在问题或偏见时，据此研发的刑事司法人工智能应用便会放大这一效应。

最后，刑事案件智能辅助办案系统可能会对审级制度形成严重冲击，损害当事人的上诉权。审级制度存在的重要价值在于赋予当事人救济机会，通过程序吸收当事人的不满，达到消弭纠纷的目的。但刑事案件智能辅助办案系统被应用于上诉机制后，无论上下级法院对刑事司法人工智能的辅助工具属性如何强调，只要上下级法院运用相同的刑事司法人工智能应用处理同一案件，就会出现完全相同的事实认定或者法律适用结论，这无疑会削弱审级制度的意义和价值。

（二）存在的正当性隐忧

在刑事司法人工智能的基本要素中，数据垄断、黑箱算法及应用场景缺乏规制，充分表明，我们的刑事司法人工智能基础薄弱。当上述不足集合于具体的刑事司法人工智能应用时，会对既有的刑事诉讼法律关系、刑事诉讼结构、刑事诉讼原则造成冲击。

首先，刑事案件智能辅助办案系统的应用将强化公检法三机关之间“配合有余，制约不足”的关系。宪法、刑事诉讼法确立了公检法三机关“分工负责、互相配合、互相制约”原则，但实践中存在“配合有余，制约不足”现象。如果运用相同的刑事司法人工智能系统办理案件，可能会使侦控审更加流水线化。既有的刑事案件智能辅助办案系统以统一证据标准为主轴进行建构，力图通过引入人工智能，实现对刑事证据的统一指引、校验、提示、把关和监督。这种思路无疑会提高办案效率，但会引发刑事诉讼流程异化，强化后续刑事诉讼流程对前续阶段的机械性认同，使三机关办案思路同质化，分工、制约关系进一步受阻。

其次，刑事案件智能辅助办案系统的应用可能会加剧本已失衡的控辩关系、辩审关系。实现控辩平衡是诉讼现代化的一项重要指标，但囿于种种限制，控辩力量对比很难达致平衡。在人工智能加持下，本已处于优势地位的侦控机关在数据获取方面天然优位，如其继续不加限制地利用数据优势和黑箱算法，将使得本已处于弱势地位的辩方更难进行有效抗辩。另外，将刑事案件智能辅助办案系统应用于事实认定阶段和量刑阶段，会导致决策权让渡，但由于辩护方对人工智能是否参与以及

^① 参见 Melissa Hamilton, “The Biased Algorithm: Evidence of Disparate Impact on Hispanics,” *American Criminal Law Review*, vol. 56, no. 4, 2019, pp. 1553-1554.

发挥作用程度大小并不知晓，当对结果有异议时，会不可避免地归因于审判人员，加剧辩审冲突。

再次，刑事案件智能辅助办案系统的透明度有待提升。数据、算法是刑事司法人工智能的核心要素，但数据垄断和算法黑箱的存在，导致刑事司法人工智能缺乏透明度，使被追诉人处于数据不对称地位和算法信息劣势。缺乏透明度的刑事司法人工智能应用不仅剥夺被追诉人的知情权、减损其抗辩权，还使得“研究人员和外部专家难以评估和审核算法，从而难以测试算法的准确性和偏差值”。^① 由此会带来对刑事司法人工智能公信力的疑虑，最终影响应用实效。

最后，难以建立针对刑事司法人工智能的监督与问责机制。司法责任制被视为本轮司法改革的“牛鼻子”，^② 完善司法责任制，是建立权责统一、权责明晰、权力制约的司法权运行机制的关键。“让审理者裁判，由裁判者负责”，既是改革的共识，也是改革的努力方向。但当下，如果因刑事司法人工智能应用造成冤错案件，监督和问责将变得困难重重。当案件裁决者由单一化走向多元化和复数化时，审理者的角色将不再确定、明晰，导致难以有效问责，会使得正在着力推进的司法责任制改革难以获得真正落实。

（三）潜在的伦理性风险

刑事司法人工智能是一把双刃剑，其“有可能提高刑事司法系统的审判准确性，并降低人为错误和偏见的风险，但它们也有可能强化或加剧现有的偏见，并破坏基本的司法正义原则”。^③ 刑事案件智能辅助办案系统在提高诉讼效率的同时，也对刑事司法领域的伦理范式、正义观念造成冲击和挑战。

刑事司法人工智能应用过程中，由人工智能自动生成判决、根据大数据矫正法律决定的偏差，特别是让人工智能参与事实认定导致决策让渡，使得法官定位发生极大动摇，甚至造成审判系统乃至司法权的全面解构，从而在事实上形成审判主体的双重结构。^④ 并且，随着机器学习的迭代升级，机器正从被动工具、辅助者转变为具有自主意识、自主决策能力的能动者。就刑事案件智能辅助办案系统而言，其具有的自动生成判决、裁判偏离度预警等功能，已经具有自动化决策功能，自主性和能动性的具备使其不再只是法官的辅助工具，而成为法官事实上的代理者甚至替

① Danielle Kehl et al. , “Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing,” *Berkman Klein Center for Internet & Society*, November 28, 2017.

② 参见习近平：《习近平谈治国理政》第2卷，北京：外文出版社，2017年，第131页。

③ Danielle Kehl et al. , “Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing.”

④ 参见季卫东：《人工智能时代的司法权之变》，《东方法学》2018年第1期。

代者，对既有刑事司法领域伦理范式构成重大挑战。^①

刑事司法人工智能中算法决策的深度应用，还对既有的司法正义结构^②形成冲击。司法改革中，刑事司法人工智能的限度并没有被充分考量，算法决策已经被应用于事实认定甚至复杂案件决策中，这实质上是对算法决策的过度迷信。既有的刑事司法人工智能设计认为利用海量数据和算法，并辅之以超强算力，就可以破解刑事诉讼中的难题，防范冤假错案，促进司法公正。^③但问题在于，司法正义特别是实质正义难以通过具体的算法实现，毕竟“司法人工智能不会计算情感，不能与人进行心灵互动，自然不可能有人文关怀，难以根据具体场景灵活地、创造性地维护正义价值”。^④但司法实践中，诸如犯罪预测、风险评估工具在内的刑事司法人工智能应用，已经在将公平问题算法化。问题在于，刑事司法人工智能中的算法决策依据的是通过数据、算法、建模构建的“计算正义”，而非人本、人文层面的正义。

刑事司法人工智能面临的合法性挑战、存在的正当性隐忧以及潜在的伦理性风险表明，刑事司法人工智能亟待规制。但同时，上述问题的复杂性也表明，规制难度巨大。全球刑事司法人工智能改革的代表性国家都没有形成体系化的监管制度，这既有刑事司法人工智能技术发展不成熟的原因，也有政策制定者无法全面认知刑事司法人工智能风险的因素。但在我国刑事司法人工智能应用已经获得长足发展、问题已经充分显现时，跨越技术规制的“科林格里奇困境”，已经迫在眉睫。

三、刑事司法人工智能的规制原理

虽然域外国家刑事司法人工智能监管领域的严格规制模式和渐进规制模式对我国具有一定启示意义，但由于我国刑事司法的独特诉讼文化和诉讼构造，加之我国处于全球刑事司法人工智能的无人区，要求通过自身探索，创制契合我国刑事司法正义体系的规制方案。

顺应新兴科技赋能刑事司法的发展趋势，以及参与全球人工智能规制竞争的宏观背景，刑事司法人工智能规制方案的设定，应当以维护司法正义为前提，通过技

① 参见腾讯研究院等：《人工智能：国家人工智能战略行动抓手》，第 290 页。

② 通说认为，司法正义是实质正义与形式正义的有机统一。实质正义是法律制度的基本价值，是司法追求的最终目标；形式正义是实质正义的基础保障，是实现正义的前提和基础。

③ 参见崔亚东：《人工智能与司法现代化》，第 74—75 页。

④ 马长山：《司法人工智能的重塑效应及其限度》，《法学研究》2020 年第 4 期。

术创新实现对人类赋权、对产业赋能。由此，能够平衡各方利益，实现激励发展与合理规制相协调的包容性规制^①就成为最佳选择。包容性规制鼓励技术创新，但并非片面的发展优位，而是立足于监管对象刑事司法人工智能的司法属性，遵循司法规律进行合理规制。由此，维护司法正义、保障基本权利，补足既往基本权利保障不足的缺环^②就成为包容性规制的出发点和着力点。

（一）技术赋权理念：以人为本与权利保障的统一

由于人工智能在相当长的时期内发展相对缓慢，人工智能应用领域的权利保护问题并未引起重视。在新一轮人工智能快速发展的背景下，人工智能带来的技术侵权问题愈发凸显。对于人工智能应用、发展应当践行何种价值观，各方思路逐渐清晰、一致，人工智能以人为本、^③以人为核心的人工智能、^④推进以人为核心的发展路径^⑤等理念逐渐为人接受。2019年二十国集团部长级会议通过的《G20人工智能原则》明确提出以人为本的人工智能原则，标志着以人为本的人工智能发展理念已经成为国际社会的普遍共识。

人工智能以人为本强调人的核心价值，与我国宪法、刑事诉讼法规定的“尊重和保障人权”理念高度契合，二者都强调对“人”本质的尊重。但在刑事司法人工智能领域，技术赋能凸显，导致以人为本理念并未实现与刑事司法保障人权理念的有机统一。刑事司法人工智能可能存在的负面效应，会对刑事司法权利造成严重威胁。并且在具体应用实践中，当解决案多人少、提高效率成为刑事司法人工智能追求的主要目标时，无论是人工智能业界倡导的以人为本，还是刑事司法追求的人权保障，都将被遮蔽。基于此，实现两者融合的关键在于，将人工智能业界推崇的以人为本、增进人类福祉与刑事司法追求的保障人权有机融合，化约为简洁的技术赋

-
- ① 李克强在2017年夏季达沃斯论坛致辞中提出，对新产业、新业态、新模式，如电子商务、移动支付，实行包容审慎监管方式，促进了其健康发展。参见李克强：《在第十一届夏季达沃斯论坛开幕式上的致辞》，《人民日报》2017年6月28日，第3版。
- ② 国家《新一代人工智能治理原则》强调和谐友好、公平公正、包容共享、尊重隐私、安全可控、共担责任、开放协作、敏捷治理等八项原则，尽管突出对于隐私权利的尊重，但并未将尊重基本权利作为基本原则予以纳入。
- ③ 参见 Will Knight, “Put Humans at the Center of AI,” *MIT Technology Review*, vol. 120, no. 6, 2017.
- ④ 参见 Fei-Fei Li, “How to Make A.I. Human-Friendly,” *The New York Times*, Mar. 8, 2018, A27.
- ⑤ 参见 European Political Strategy Centre, “The Age of Artificial Intelligence: Towards a European Strategy for Human-Centric Machines,” *EPSC Strategic Notes*, vol. 29, March 27, 2018.

权理念。^①即通过具体刑事司法人工智能应用，在技术赋权理念重塑下，刑事司法人工智能的研发、应用和评估以是否有助于保障刑事司法权利，是否有助于发展刑事司法权利，是否有助于创制刑事司法权利为判断基准。^②

在刑事司法人工智能规制领域，技术赋权理念坚持问题导向，制约技术赋能带来的国家基础性权力的扩张给基本权利保障带来的威胁；^③技术赋权理念坚持目标导向，引导刑事司法人工智能的研发，以提高刑事司法领域的权利保障水平为依归；技术赋权理念坚持结果导向，以基本权利是否得到保障、提高，作为评判具体刑事司法人工智能应用的尺度。由此，技术赋权理念既可以为刑事司法人工智能的研发提供事前规制，又可以为其实实施提供过程指引，还可以提供事后评价标准。

技术赋权理念贯通人工智能以人为本与刑事司法保障人权理念，实现人工智能与刑事司法的有效连接，这既是对技术赋能被过度强调的纠偏，又是强化刑事司法人工智能应用中权利保障的必然选择。更为重要的是，技术赋权理念具有契合现实的可行性，其遵循刑事司法规律，顺应人工智能业界“科技向善”的使命愿景，在将刑事司法保障人权理念与人工智能业界以人为本理念对接的同时，成功实现两种理念的融合，提升为数字时代具有可操作性的监管理念。同时，技术赋权理念不仅可以规制当下以人工智能、大数据为代表的前沿科技，还具有规制未来新兴技术的前瞻性和包容性。

（二）技术正当程序：一种新的程序正义观

刑事司法人工智能应用是技术在刑事司法程序的推进与展开，作为刑事司法人

① 郑永年运用技术赋权来分析互联网是如何影响中国国家—社会关系，来解释中国的互联网政治。本文将技术赋权理念应用于对刑事司法人工智能规制的分析中。参见郑永年：《技术赋权：中国的互联网、国家与社会》，邱道隆译，北京：东方出版社，2014年，第1页。

② 以创制权利为例，欧盟《通用数据保护条例》因应科技发展创制出数据转移权、自动决策的可解释权，如将自动决策的可解释权推广应用于刑事司法人工智能领域，就可以解释为权利人有权要求刑事司法人工智能应用主体就算法决策作出解释。参见 Bryce Goodman and Seth Flaxman, “EU Regulations on Algorithmic Decision-Making and a ‘Right to Explanation’,” *AI Magazine*, vol. 38, no. 3, 2017, pp. 50-57; Gianclaudio Malgieri, “Automated Decision-Making in the EU Member States: The Right to Explanation and other ‘Suitable Safeguards’ in the National Legislations,” *Computer Law & Security Review*, vol. 35, no. 5, 2019, pp. 23-25.

③ 基础性权力是指国家能实际穿透市民社会并依靠后勤支持在其统治的疆域内实施其政治决策的能力，它通过对社会生活的渗透，通过对社会行动的协调和服务来实现治理目的。国家能够运用的技术越是发达，国家的基础穿透能力就越强。参见刘昶：《迈克尔·曼论国家自主性权力》，《上海行政学院学报》2016年第1期。

工智能核心的算法甚至被通俗地视为“解决问题的进程”,^① 自带程序属性的刑事司法人工智能,更需要进行过程控制。但问题在于,现代刑事诉讼法的正当程序理论产生时并未发生人工智能技术的司法应用问题,既有的刑事诉讼法律制度能否有效规制新兴的刑事司法人工智能,值得深入研究。

传统正当程序理论强调裁判者的独立、中立、公开、参与等核心要素,^② 力图通过为诉讼程序设置标准来实现程序正当化。在人工智能时代,算法主导、实行自动决策的刑事司法人工智能依照传统正当程序标准适用时,条件发生了重大变化,不仅侵蚀了传统正当程序推崇的获得通知、听证的权利,而且危及参与、透明的正当程序要素。^③ 不仅如此,刑事司法人工智能所附带的算法黑箱及问责问题,传统正当程序理论对此也无能为力。凡此种种,传统正当程序理论正面临严峻挑战。正是在此背景下,算法正当程序、程序性数据正当程序等概念频频出现,技术正当程序理论应运而生。^④

技术正当程序强调透明、准确、可问责、参与等核心要素,^⑤ 其核心使命在于克服传统正当程序理论对人工智能背景下司法运作的解释力失灵问题。技术正当程序并非否定传统正当程序理论,而是通过化解传统正当程序理论内含的独立、中立、公开、可问责等要素所面临的正义风险,发挥对于正义实现的促进价值,^⑥ 通过数据公开、算法透明和建立问责机制,与传统正当程序理论相互支撑,共同实现人工智能背景下的公平正义目标。

1. 推进司法数据的公开与算法的透明

在传统正当程序理论中,透明主要指贯彻公开审判原则,公开审判过程。在刑事司法人工智能时代,在技术正当程序理论中,透明不仅包括程序公开,更涵盖司法数据的公开以及算法的开源,更多地指向算法黑箱问题。算法不透明普遍

① 瑟格·阿比特博、吉尔·多维克:《算法小时代:从数学到生活的历变》,任轶译,北京:人民邮电出版社,2017年,第3页。

② 参见季卫东:《法律程序的意义》,北京:中国法制出版社,2012年,第18页。

③ 参见 Danielle Keats Citron, “Technological Due Process,” *Washington University Law Review*, vol. 85, no. 6, 2017, p. 1249.

④ 参见 Danielle Keats Citron, “Technological Due Process,” p. 1249.

⑤ 参见 Danielle Keats Citron and Frank Pasquale, “The Scored Society: Due Process for Automated Predictions,” *Washington University Law Review*, vol. 89, no. 1, 2014, p. 20.

⑥ 在技术正当程序规制刑事司法人工智能的框架下,透明、可问责、监督以及公平成为规制的切入点。参见 John Lightbourne, “Damned Lies & Criminal Sentencing Using Evidence-Based Tools,” *Duke Law & Technology Review*, vol. 15, no. 1, 2017, p. 343; Katherine Freeman, “Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in *State v. Loomis*,” *North Carolina Journal of Law & Technology*, vol. 18, online edition, 2016, p. 78.

为人诟病,^①刑事司法人工智能中的算法黑箱更是成为众矢之的,以商业秘密为由拒绝公开算法已经难以成为正当借口。^②

有学者认为,刑事司法人工智能中的算法高度复杂,非常人所能理解,即便公开也并无多大价值。亦有学者认为,对于正当程序而言,代码开源既非必要又不充分,因此透明并非必要要素。^③但问题在于,不透明的刑事司法人工智能直接侵害被追诉人的知情权外,还导致难以对算法进行审核,无法测试算法的准确性,进而导致无法对刑事司法人工智能的裁判结果有效行使抗辩权。随着研究的深入,开源算法为越来越多的人所认同,有观点甚至主张公开计算机程序检测科学证据的源代码。^④毕竟更大的透明度可以帮助外部研究人员进行研判,可以帮助增加对这些系统的总体了解,知悉运作过程及权衡的各种因素。^⑤

当下我国的刑事司法人工智能应用实践中,传统正当程序理论中的诉讼公开,技术正当程序理论中的司法数据公开及算法开源叠加在一起,使程序公开面临的问题更为复杂。深入推进刑事诉讼程序改革虽然有助于实现程序公开,但如果不及时解决数据公开和算法黑箱问题,程序公开将面临严峻挑战。为此,原有的程序公开原则需要增加新的理解变量,需要通过完善司法数据公开平台建设进一步推进刑事司法数据的公开,更需要通过推动刑事司法人工智能研发机构开源算法以实现算法的透明化。^⑥

2. 通过推进算法可解释性构建问责机制

尽管透明度对于刑事司法人工智能至关重要,但公开数据和算法只是开始,亟须在此基础上建立刑事司法人工智能问责机制。毕竟,对于海量数据,缺乏专业分析工具,普通人难以获知其中的规律、难以预测未来的趋势。对于算法,非专业人士更是难以理解。尽管刑事司法人工智能的研发主体宣称基于正当目的设计人工智能,并确保将其建立在科学的基础之上,但如果缺乏相关的监督与问责机制,并不

① 参见弗兰克·帕斯奎尔:《黑箱社会:控制金钱和信息的数据法则》,赵亚男译,北京:中信出版集团,2015年,第1页。

② 参见 Rebecca Wexler, "Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System," *Stanford Law Review*, vol. 70, no. 5, 2018, p. 1343.

③ 参见 Joshua A. Kroll et al., "Accountable Algorithms," *University of Pennsylvania Law Review*, vol. 165, no. 3, 2017, pp. 657-660.

④ 参见 Christian Chessman, "A 'Source' of Error: Computer Code, Criminal Defendants, and the Constitution," *California Law Review*, vol. 105, no. 1, 2017, pp. 199-221.

⑤ 参见 Danielle Kehl et al., "Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing."

⑥ 刑事司法人工智能应用中的算法应当以公开为原则,当与国家安全、社会秩序冲突时可以例外地不予公开。关于算法透明与国家安全、社会秩序冲突及私主体权利冲突的论述,参见沈伟伟:《算法透明原则的迷思》,《环球法律评论》2019年第6期。

能从根本上解决问题。

传统正当程序理论中，诉讼各方参与刑事司法裁判过程，既能够监督裁判过程，也能够对最终的裁判结果发起救济。但在运用刑事司法人工智能的场景中，相关诉讼主体却难以对刑事司法人工智能应用进行有效监督。基于此，技术正当程序特别强调刑事司法人工智能的可问责性。^① 建立人工智能可问责机制有多种方式，^② 在法律领域，要求人工智能研发者对算法作出必要解释作为一种有效手段被特别强调，^③ “在司法裁决中，凡涉及自主研制系统，都应提供一个有说服力的解释，并由一个有能力胜任的人员进行审计”^④ 这一原则已经为越来越多的人认可、接受。

要求刑事司法人工智能研发主体承担解释责任，具有如下积极意义：（1）有助于推动自动决策领域可解释权的建立，可解释性与公民的可解释权系一体两面，强调研发、应用主体的可解释性义务是在反面推动赋予公民可解释权；（2）有助于约束刑事司法人工智能的研发、应用主体，要求其在研发、应用时承担解释的义务，能够防范可能的设计或者应用缺陷；（3）有助于刑事司法人工智能的具体应用主体如法官准确理解、接纳和应用刑事司法人工智能；（4）有助于增强民众对刑事司法的信任，尽管民众未必理解刑事司法人工智能的运作原理，但通过对刑事司法人工智能的运作特别是算法作出解释，能够展现研发、应用主体的自信与坦诚，有助于获取民众的信任；（5）有助于外界对刑事司法人工智能进行验证、监督。刑事司法人工智能的研发、应用主体对具体刑事司法人工智能应用作出解释，能够促进新闻媒体、科技公司对该应用进行验证，科技公司的介入还会促进该刑事司法人工智能应用的迭代升级。由此，当人工智能研发主体不尽释明责任时，应当承担相应的法律责任。

技术赋权理念和技术正当程序的双重规制体系并非是对传统程序理论的否定，而是对既有刑事诉讼理论的继承、发展和创新。我国在面对新兴科技给刑事司法带来挑战的同时，还需要解决传统程序的正当化问题，只有实现传统程序理论与新技术背景下程序理论的有机统一，才能真正推动司法正义的实现。

① 参见 Danielle Keats Citron, “Technological Due Process,” p. 1249; Danielle Keats Citron and Frank Pasquale, “The Scored Society: Due Process for Automated Predictions,” p. 18.

② 参见 Joshua A. Kroll et al., “Accountable Algorithms,” p. 695.

③ 参见 Finale Doshi-Velez et al., “Accountability of AI Under the Law: The Role of Explanation,” SSRN Electronic Journal, November 27, 2017.

④ Marcin Garbowski, “A Critical Analysis of the Asilomar AI Principles,” *Zeszyty Naukowe Politechniki Slaskiej*, vol. 115, 2018, p. 51.

四、刑事司法人工智能的规制路径

基于激励发展与合理监管相协调的发展目标,刑事司法人工智能具体规制方案的设定应当践行包容性规制理论,有所为有所不为,根据《新一代人工智能发展规划》提出的“建立健全公开透明的人工智能监管体系,实行设计问责和应用监督并重的双层监管结构,实现对人工智能算法设计、产品开发和成果应用等的全流程监管”,结合发展实际与未来可能,实行区别对待的分类监管策略。

第一,设定刑事司法人工智能的应用场景。人工智能赋能传统产业带来的最大变化是人工智能接管人类的重复性工作,刑事司法领域中的重复性、可替代性工作均可由人工智能承担,例如,法律检索、法律咨询、庭审录入等工作。由刑事司法人工智能应用从事可重复性工作,不仅会大幅减轻司法人员的工作量,还会促进司法人员创造力的释放。而且,刑事司法人工智能应用并非只是单纯替代司法人员从事重复性、机械化工作,其还可以在某些方面实现创造性升级。例如,利用我国在自然语言处理领域的优势,着力研发可以为当事人提供法律咨询、法律检索的高水平问答机器人,不仅可以改变当下低质低效的状况,还可以有效解决值班律师不足的问题,促进司法的可及性,通过科技实现普惠司法。再如,可以将风险评估工具引入审查逮捕及羁押必要性审查阶段,替代取保候审保证金制度,让被追诉人享受科技红利,避免因经济困难而被剥夺取保候审的机会。^①

第二,建立刑事司法人工智能应用负面清单。刑事司法人工智能技术发展尚未成熟及其存在的风险和不确定性,促使我们应当研判其适用的边界和限度。首先,从技术层面来看,当下人工智能处于弱人工智能时代,强人工智能在短期内难以实现,使人工智能赋能的领域和边界受到限制。其次,从价值层面来看,鉴于人工智能在创意、情感和裁判领域的天然缺陷,要实现情理法高度融合的裁判正义,具有技术上难以克服的障碍。最后,从司法实践层面来看,既往刑事司法人工智能应用过程中,过于强调技术对于司法的引领作用,对于司法规律把握不足。

基于此,应根据刑事司法的价值追求和技术发展的现实水平为刑事司法人工智能设定必要的限度。凡是可能导致不可欲的刑事司法风险,可能造成当事人权利减损,可能冲击既有刑事司法正义结构的人工智能应用都应当列入负面清单。例如,应当禁止刑事司法人工智能进行自主决策、禁止其从事证据审查判断等事实认定工作。当然,负面清单要因应技术发展、法律变动进行动态调整。

^① 如果只是以既有的逮捕数据设计人工智能应用,将难以降低居高不下的逮捕率。刑事司法人工智能设计需要警惕运用既有数据进行机器学习存在的弊端,避免底层数据可能存在的偏见而导致的“技术侵权”现象发生。

第三，完善刑事司法人工智能应用的审核机制和责任追究机制。由于刑事司法人工智能应用关涉刑事司法正义、基本权利，同时又涉及技术创新和研发主体的经济利益，建立差异化的审核机制，能够有效平衡各方利益。首先，应当区分刑事司法人工智能的私人使用和公共使用。对刑事司法人工智能的私人使用以不侵犯公共利益、私人权利为监管原则，但严格防范资本裹挟下科技对司法可能的侵蚀。^①其次，对刑事司法人工智能的公共使用进一步采取差异化规制策略，对智能辅助决策应用系统建立专门的审核、认证机制，以是否实现技术赋权理念为判断基准，进行全面事前审查，决定是否颁发证书；对优化管理、提高效率的智能管理应用系统和提升服务质量的诉讼服务应用系统，则可以采取适度宽松的备案制。

在建构事先约束的审核机制的同时，还需要完善刑事司法人工智能的责任追究机制，以实现事后惩戒与事先约束的协同作用。当然，法律责任的设定，应当合理适度，不能因为追责机制的建立而束缚研发人员的创新精神，阻碍刑事司法人工智能技术的创新发展，从而违背包容性规制原理。基于此，刑事司法人工智能的责任设定，应当以当事人权利实质受损为前提，区分研发、审核、应用主体并采取差异化对待策略。权利遭受侵害的当事人，以违反技术正当程序为由寻求司法救济，作为应用主体的刑事司法机关应当为其提供充分救济渠道；当前述技术侵权确定后，刑事司法机关可以追究研发主体和审核主体的责任，但后者的责任承担以存在故意或重大过失为限。

对刑事司法人工智能进行规制的根本目的在于促进司法公正与效率的双重提升，这一目标的实现也有赖于刑事诉讼能够因应技术发展进行流程改造。在遵循司法固有规律的前提下，探索刑事司法人工智能不同应用场景下当事人诉讼权利的实现方式，真正通过技术正当程序落实技术赋权的理念。

结 语

刑事司法人工智能应用的经验表明，以人工智能为代表的新兴科技赋能司法领域已经是大势所趋。但科技融入司法可能引发的道德伦理、正义观念、法律适用风险已经凸显，并对既有法律制度和监管模式形成冲击。这些都是新时代法治建设需要面对的重大命题。

人类科技史上的许多重大进步都曾促进司法领域的深刻变革。司法证明方式从神证到人证、从人证到物证的两次重大转变，^②都是科技发展助推的结果。科技进

① 由于法律科技巨大的市场前景，阿里巴巴、腾讯等互联网企业均已开始布局，法律科技的快速发展也催生了大量“技术驱动法律”的初创企业。

② 参见何家弘：《司法证明方法和推定规则》，北京：法律出版社，2018年，第36页。

步使得人类逐渐摆脱蒙昧，祛除了司法裁判的神明化，技术的快速发展使得长期处于随机变化和分散的物证在司法活动中作用凸显。法庭科学以及各种人身识别技术的兴起，为 19 世纪科学证明方法的发展奠定了坚实基础。^① 20 世纪末以来新证据学的兴起，证据科学时代的来临，更是与交叉学科、新兴科技的快速发展直接相关。既往刑事司法领域的深刻变革，都是人类主动应用新技术的结果，体现的是人类主导下的技术工具化。

当下以大数据、人工智能为代表的新一轮科技革命，呈现出鲜明的特点，不再是既往技术工具化的线性演进，技术本身开始展现出极强的智能性和自主性，这在相当程度上改变了司法领域中的人与机器的关系。在人机关系新型态下，机器不再只是单纯的工具，数据驱动下的智能机器开始具备自动化裁判能力。这一巨大转变对既有的司法结构和司法治理模式形成巨大冲击，并对既有司法制度产生重塑效应，逐渐形成人机协同的司法治理新模式。^② 在技术乐观主义者看来，这是在利用机器智能培育现代司法体制的新动能，代表着司法的未来发展方向。悲观主义者则认为，人工智能在司法领域中的应用会导致技术对人的宰制。

面对司法领域人机关系的新型态，应摒弃司法科技领域中流行的技术决定论，充分研判人机关系变革可能引发的法律、伦理及社会问题，遵循刑事司法人工智能的司法属性，坚持以人为本的司法正义理念，践行技术赋权和技术正当程序，实行包容性规制。通过激励发展与合理规制的协调，为全球刑事司法治理提供中国经验，提升我国在法律科技领域的国际话语权和规则制定权。^③

〔责任编辑：李树民〕

① 参见何家弘：《司法证明方法和推定规则》，第 43 页。

② 党的十九届四中全会提出“坚持和完善共建共治共享的社会治理制度”，“打造数据驱动、人机协同、跨界融合、共创分享的智能化治理新模式”成为其中的重要组成部分。参见郭声琨：《坚持和完善共建共治共享的社会治理制度》，《人民日报》2019 年 11 月 28 日，第 6 版。

③ 习近平在阐述网络信息技术、区块链应用等新兴科技时强调提升“国际话语权和规则制定权”。参见习近平：《加快推进网络信息技术自主创新 朝着建设网络强国目标不懈努力》，《人民日报》2016 年 10 月 10 日，第 1 版；习近平：《把区块链作为核心技术自主创新重要突破口 加快推动区块链技术和产业创新发展》，《人民日报》2019 年 10 月 26 日，第 1 版。

that equate the universality of philosophy with Western philosophy; besides, normativity cannot be understood as specific censorship rules. The universality of philosophy requires a sense of boundaries and a normative attitude that will ensure that people from different contexts can have dialogue on the basis of rational ability and construct a space for dialogue and understanding. The concern for diverse objects based on the richness of experience gives philosophical concepts and arguments pluralist differences that encompass cultural differences and autonomy in terms of level of expression. All in all, it is only on the basis of the universality of philosophy that we can understand the normative requirements of philosophical activities and the diversity of cross-cultural philosophical reflection.

Inclusive Regulation of Artificial Intelligence in Criminal Justice *Li Xunhu* • 42 •

Although the artificial intelligence (AI) of China's criminal justice leads the world, problems such as data monopoly, algorithmic black boxes, and arbitrary application scenarios have been exposed. In order to resolve these problems in a scientific way as well as the legality challenges, legitimacy concerns and ethical risks underlying them, we need to overcome traditional way of thinking, innovate our regulatory models, introduce the concept of technology empowerment and technological due process, and implement inclusive regulation of AI in criminal justice so as to achieve coordination between development and regulation. The deep integration of human rights protection in criminal justice and human-centered AI form the concept of technology empowerment, which can provide macro guidance at the conceptual level for the research and application of artificial intelligence in criminal justice. Technological due process, which is based on the traditional due process, can regulate the process of AI applications for the openness of judicial data, enhance the transparency of algorithms, and promote the establishment of an AI accountability mechanism in criminal justice. The application of data-driven judicial AI will influence the judicial structure and governance mode and have the effect of reshaping judicial operations, gradually forming a new model of judicial governance with man-machine collaboration. In the future, we should uphold the concept of human-centeredness and the idea of technological empowerment, reshape our value orientation and code of conduct under the new configuration of man-machine relations, and pursue the values of legal fairness and justice, in hopes of building a new pattern of man-machine collaborative judicial governance.

Big Data Technologies and the Modern Transformation of Traditional Philology

Liu Shi and Li Feiyue • 63 •

Big data technologies have triggered innovations in the production methods and structural transformation of traditional philology and have expanded its accessibility. The fragmentation, standardization, structuration and visualization of traditional philology constitute the “macro-texts” and “hypertexts” consisting of the various textual collections and databases, which have promoted the connections between different kinds of literature and the rediscovery of knowledge. The application of online analysis, bibliometrics, thematic modeling and other textual information

• 205 •