

# 司法人工智能的重塑效应及其限度

马长山\*

---

**内容提要：**人工智能技术的司法应用，开启了从“接近正义”迈向“可视正义”的历史进程，引发了司法运行机制的深刻变革，通过司法过程场景化、司法规则代码化、司法决策建模化、司法服务管理智慧化重塑着司法过程，大大提升了智慧司法的社会效能。但司法人工智能亦有其内在限度。算法决策要达到绝对的客观和精准并非易事，面对复杂疑难案件，其能够促进形式正义，却很难实现实质正义。发挥人工智能对司法的优化重塑效应，应与对人工智能的风险防范同步进行。需针对算法决策设立相应的规制和救济机制，注入公开、公平和责任理念，进而兴利除弊，促进科技向善和打造良性发展的智慧司法，更好地实现“可视正义”。

**关键词：**司法人工智能 可视正义 数字正义 智慧司法

---

随着网络化、数字化、智能化的飞速融合发展，人类逐渐步入智慧时代和数字社会。近年来，我国的智慧政务、智慧司法、智慧治理等重大建设方案纷纷落地。在这些方案的实施过程中，司法人工智能无疑将会扮演关键角色并承担重要使命，〔1〕同时也会面临诸多问题和挑战。理性审视司法人工智能的应用空间与限度，由此成为一项重要的理论课题和实践任务。

## 一、从“接近正义”迈向“可视正义”

设置司法权的根本意义在于处理纠纷、实施救济、弘扬正义，但要让人们在每一个案件中切身体验正义，却不那么容易，这也正是卡夫卡“法律之门”的深层隐喻所在。人们曾对“法律之门”做出很多不同解读，但有两点是明确的：其一，“法律之门”是“供给”正义的地方，它庄严尊贵却又壁垒森严；其二，人们可以走近“法律之门”，但难以迈入门

---

\* 华东政法大学教授。

〔1〕 在学术界，人们曾对 AI 法律和 法律 AI 进行过区分和讨论。当今社会，网络技术、数字技术和人工智能技术日渐呈现交融发展之势，三者相互支撑和彼此赋能。因此，本文所称“司法人工智能”是一个涵摄面更广、包容性更强的概念，主要指司法领域中以网络化、数字化和智能化融合驱动为基础的人工智能技术应用。

内,更难观测门内的设置、工序流程和决策机制。这很大程度上要归因于物理时空中的资源限制、区间阻隔和机制障碍。特别是,由于交通不便、诉讼费用高昂、诉讼过程繁琐漫长等原因,人们在寻求司法救济时经常遇到挫折、面临困难,以至于20世纪中叶以来,“司法危机”在世界范围内大面积出现。为摆脱这一困境,人们发起了旨在为当事人接近法院、实现诉权提供实质保障的“接近正义”运动。该运动的基础和前提是国家财力和司法改革,价值导向是平等和效率,目标是实现分配正义。经过“三波”重要努力,<sup>[2]</sup>“接近正义”运动取得了一定效果。然而,由于福利国家的不可持续性以及司法改革的制度性瓶颈,“接近正义”运动需要付出较高成本和代价,也面临诸多难题。这些局限在传统的物理时空和策略路径下很难克服。

信息革命的到来,给“接近正义”运动的整体转型提供了巨大动力和难得契机。算法的应用不仅提升了司法的纠纷处理能力,降低了解决纠纷的成本,而且在很大程度上提高了纠纷解决的自动化程度,“为更加接近正义奠定了现实基础”。<sup>[3]</sup>比如,从ADR迈向ODR,增加了司法程序的公开透明度,使人们看到了“数字正义”的曙光。近年来,我国司法机关也利用这些新技术进行了大量改革探索,致力于构建“网络化”“阳光化”“智能化”的新时代司法样态,实现了对物理时空资源和要素的优化重组。通过技术“克服或至少是瓦解正义之路上的一些障碍”,<sup>[4]</sup>可以消解因物理时空条件局限而导致的正义实现难题,使物理意义上的“接近正义”迈向数字意义上的“可视正义”。

其一,平台化的分享可视。新兴技术革命颠覆了传统的工商业态和经营模式,人类从此进入平台经济时代,平台模式也迅速地从工商业向社会领域延展,其突出特点和优势在于:线上线下交融、多主体共享、生态化运行、覆盖性整合以及价值链创造。这就突破了以往物理时空的商业逻辑,颠覆了单一线性的交易形式,开启了广场化的全景互动。智慧司法建设同样也经历了这种技术赋能过程,深刻重塑着司法运行模式,具体表现为:在案件办理上,支持全流程在线审理,办案全过程智能辅助,审理信息全方位在线公开;在监督管理上,实现重点案件自动化识别、标签化处理、节点化控制,建立智能化、自动化、精准化监管机制;在诉讼服务上,形成多功能、集成性、智能化、线上线下融合的一站式诉讼服务模式;在平台载体上,实现内部平台整合对接,外部数据互联互通,形成系统集成、集约高效、信息共享的平台建设模式。<sup>[5]</sup>司法运行的这种平台化转向,能够“让数据发声,使数据关联,用大数据思维发现审判规律、提高审判质效”,<sup>[6]</sup>也能够在微观(司法机关内部)、中观(司法机关与当事人和相关机构)和宏观(司法机关与全社会)三个层面,分别实现不同程度的数据分享和在线互动,覆盖几乎所有司法业务,形成司法运行的

[2] 即从为当事人提供直接的物质帮助和法律援助,到推进扩散利益保护的公益诉讼和司法改革,再到提供诉讼外的替代性纠纷解决方案(ADR)这“三波”运动。参见[意]卡佩莱蒂编:《福利国家与接近正义》,刘俊祥等译,法律出版社2000年版,第5页。

[3] [美]伊森·凯什、[以色列]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,法律出版社2019年版,第64页以下。

[4] 同上书,第256页。

[5] 见《为网络社会治理贡献法治智慧》,《人民法院报》2019年9月7日第2版。

[6] 李阳:《打造智能时代司法文明新坐标——人民法院智慧法院建设工作综述》,《人民法院报》2019年3月10日第7版。

“生态圈”和正义“价值链”。这样，司法运行从线下走到了线上，从封闭的、局限的“剧场”模式转向了开放的、包容的平台模式，实现全程可视。曾经神秘威严的“法律之门”被移到阳光下，不仅提升了司法效率，促进了司法公正，也便于公众监督和正义传播。

其二，超时空的场景可视。对司法程序进行高度专业化、复杂化的制度设计，目的在于阻断人为因素干扰和不公因素渗透，从而更好地保护权利和维护正义。然而，这也带来一个负面问题，即“法律之门”变成了繁琐昂贵的博弈场。“接近正义”运动的诸多举措之所以未能取得理想效果，原因在于其以物理时空的思维和方式解决物理时空的问题，无法打破“法律之门”的物理时空“围墙”。随着数字时代的到来，原来的物理“围墙”逐渐被虚拟空间击碎；原来生物形态的身体、行为、言语，原来物理形态的财物、流转及其关系等，逐渐增添了一种“无形流动”的数字形态。此时，很多旧有的习惯将被颠覆，旧有的制度将面临挑战，司法制度及其运行机制也随之发生深度变革。“在‘数字正义’理念之下，私人和社会都加强了对数字技术的采用，在法院内外共同促进‘接近正义’的实现”，尤其是人们通过设计出可以服务公众的虚拟空间，使纠纷解决克服物理条件的限制。<sup>〔7〕</sup>例如，各地开发上线的“微诉讼”“E法庭”“律师通”“执检小智”以及正在全国推广的“移动微法院”等小程序或软件系统，使当事人足不出户即可参与在线庭审、调解或执行。人脸识别、智能验证、全程视频留痕，也让司法运行更加规范透明。这样，凭借技术进步就实现了线上与线下、虚拟与现实、远程与现场的立体链接与跨界融合，甚至形成了当事人“零跑腿”的司法环境。“法律之门”也因此摆脱了物理形态的约束，开始在社会中进行超时空、无形化、分布式的前置延展，形成了灵活便捷、体验多样的场景可视状态。据此，普通民众得以随时随地进出“法律之门”，主张权利和伸张正义。

其三，全要素的数据可视。在新兴技术的驱动和支撑下，司法运行数据化、自动化趋势不断加速。一是基于智慧司法的全流程在线审理和平台化运行，会产生海量的司法数据，形成案件信息库、案件特征库、电子卷宗库、证据标准库、罪名要件库、裁判文书库、案例库、法律法规司法解释库、办案业务文件库等多类型数据库。这些数据库实质上是把复杂的司法案件转换为可以存储、计算和分析的数据，对其中涉及的人财物、社会关系和规则程序进行了全要素的模块化、可视化构建。这些“越来越多的数据将会讲述越来越多的故事”，<sup>〔8〕</sup>展示越来越丰富的正义价值和司法效果。二是司法程序凭借数字技术不断拓展和升级，比较典型的是司法区块链。“区块链极大的去中心化、不可篡改性、不可否认性、公开透明性共同促成了区块链作为信任基础设施的可行性，解决了参与者之间的共识问题。”<sup>〔9〕</sup>在电子证据认证和智能合同履行上，司法区块链可以做到电子数据及行为的全流程记录、全链路可信、全节点见证、全方位协作，从而实现全要素的数据可视。三是随着智慧司法的进一步发展，算法决策将被更多地应用于程序控制和司法裁判之中，人机互动、人脑算法和机器算法的融合也会日益加强。算法决策的优势在于自动运行、精细一致、超能高效。作为办案通用的“公共决策工具”，它能够全程回溯、数据留痕，这就比人脑决策

〔7〕 前引〔3〕，凯什等书，第30页以下。

〔8〕 [美] Phil Simon：《大数据可视化：重构智慧社会》，漆晨曦译，人民邮电出版社2015年版，第45页。

〔9〕 徐格等：《算法统治世界——智能经济的隐形秩序》，清华大学出版社2017年版，第311页。

更有利于聚集全要素,也更有利于可视化。总之,司法运行的数据化呈现,使其日益变得可代码化表达、可量化计算、可分析预测、可审计问责,从而实现了全要素的数据可视,并进入了“可视正义”时代。

## 二、司法人工智能的重塑效应

司法人工智能所带来的,已不再是过去那种外在技术装备的更新换代,而是对司法运行的内嵌式、技术性变革,对司法运行产生了重塑效应。

### (一) 司法过程场景化

古往今来,司法裁判过程都是在固定的场所——法庭中进行的。庭审活动包含着时间、地点、人物、事件等诸要素,经由法官、检察官、律师、证人、被告人、法警等角色的参与,按照形式化的规则和程序来完成。它类似于上演一幕幕戏剧的“剧场”运行状态,深嵌于特定化、场域化的物理时空之中,“法律之门”也就富有了神圣性、神秘性和仪式性。然而,人工智能的司法应用,使得这种场域化的“剧场”运行状态发生了重大转向。

#### 1. 远程临场

新兴科技和数字经济的发展,改变了社会的“时空体制”。社会亲近性与物理邻近性、社会相关性与空间邻近性开始剥离,空间位置或环境对大多社会过程而言已不再重要或不再具有决定性。时间和空间已经“脱嵌”并获得重构。<sup>[10]</sup>司法人工智能就在这种“时空体制”变革中,创新了司法运行机制。如作为全国试点推广的上海“206系统”(智能辅助办案系统),就涵盖了从刑事案件立案、侦查、审查起诉、庭审、判决,链接到刑罚执行、减刑假释、服刑人员刑满释放、回归社会等环节,“以实现人机协同、线上线下跨界融合,实现从刑事办案单一系统向刑罚执行、社会治理领域综合系统的转变”。<sup>[11]</sup>案件当事人无论身在何处,只要有在线电脑、摄像头、麦克风和可上传的诉讼文书,即可参加诉讼庭审,甚至允许当事人利用“碎片化”时间,灵活选择场所,以非同步的方式完成诉讼,实现超时空的“错时审理”。<sup>[12]</sup>这就突破了司法过程的物理场域限制,跨越了“虚拟/现实”的两重空间和人的“生物/数字”的双重属性,实现了数字身份及行为的远程临场。

#### 2. 情境互动

智慧法院建设的基本方向是网络化、阳光化、智能化。目前全国三千多家法院已基本实现网上立案、庭审直播、远程开庭、远程调解。诉讼过程已不再必须经过物理时空中的“对簿公堂”,而是以“云法庭”的视频庭审形式,完成人脸识别、证据上传、视频抗辩、电子笔录、全程留痕等一系列虚实交融的庭审活动。一次庭审可能出现多角色、多地点、多场所、多方式的同时在线和立体情境,每个诉讼当事人都是网络链接、信息传递、节点控制的参与者、分享者和承担者。这就改变了封闭单一、机械受动的“面对面”场域化庭

[10] 参见[德]哈特穆特·罗萨:《新异化的诞生——社会加速批判理论大纲》,郑作或译,上海人民出版社2018年版,第118页;[英]安东尼·吉登斯:《现代性的后果》,田禾译,译林出版社2011年版,第18页。

[11] 崔亚东:《人工智能与司法现代化》,上海人民出版社2019年版,第261页。

[12] 见《法治蓝皮书:2019年全国高级法院支持网上立案率首次达到100%》,http://www.chinapeace.gov.cn/chinapeace/c100007/2020-06/03/content\_12356666.shtml,2020年6月3日最后访问。

审过程，形成了灵活自主、多相界面、情境互动的司法场景。

### 3. 全景敞视

福柯曾对环形监狱的“全景敞视建筑”进行过深刻分析：“在环形边缘，人彻底被观看，但不能观看；在中心瞭望塔，人能观看一切，但不会被观看到”，从而形成了一种“分解观看/被观看二元统一体的机制”。<sup>[13]</sup>它通过空间阻隔、单向透明和中心瞭望的等级体系构架来完成，旨在确保权力的随时俯瞰、持续监视和自动运行，是一种物理时空的中心化、等级化、不对称的“全景敞视”。司法人工智能创造了一种与之相反的全新景象。在以往的庭审活动中，审判席处于法庭的正中位置且高出地面，诉讼两造分列两侧，略低于审判席，公诉人、律师、被告人、被害人、证人以及旁听席位等坐落有序。它无疑是中心化、场域化的，其认证、质证、抗辩等过程，都因物理条件限制而只能有限可视。不消说法庭外的庭审参与，即便是在法庭现场，也需工作人员的证据传递帮助才能做到“面对面”的认证质证。然而，网络化、数字化和智能化技术的融合发展与司法应用，打破了这种物理屏障，凭借电子文书、数据传输、即时视频等方式，诉讼参加人无需“相逢”即可“见面”，审判席也悄然无形化为一个阳光性的控制节点，进而改变了法庭的中心化、等级化设置。它与“全景敞视建筑”的目标相反，推动了中心权力的非中心化，其自动运行也并非那种俯瞰式的持续监视，而恰恰是趋向于参与各方的对称透明和平向可视，形成了全新的跨时空、无屏障、分布式、扁平化的“全景敞视”。

综上，在新的“时空体制”下，“社会关系从彼此互动的地域性关联中，从通过对不确定的时间的无限穿越而被重构的关联中‘脱离出来’”。<sup>[14]</sup>司法过程也不再局限于物理上的结构性设置和实体运行，而是从场域化走向了场景化，附加了数字化的意义展示、体验分享和“可视正义”。

#### （二）司法规则代码化

司法运行需要大量专业人员，他们是促成法律规则转化为社会秩序的行动者、构建者。人们在司法者身上寄托了不偏不倚、客观公正的浓厚期待，但他们中的每一个都具有人性，都有普通人的一面。个体的价值观、心情、精力甚至先入之见等，都会深深影响他们的自由裁量，这种来自人性自身的局限是无法克服的。立法本意、法律规则和正义价值在转化为社会秩序的过程中，难免受到办案人员一定程度的主观雕塑或续造，其结果可能是更趋于公正，也可能相反。同时，程序繁冗、案多人少、专业素质等因素也影响着司法成本和司法效率。这些一直是迈进“法律之门”和“接近正义”所必须承受的“司法代价”。

信息革命把人们带进了虚实同构的“在线生活”，形成了一套“法规的/自愿的、正式的/非正式的、国家的/超国家的、等级制的/分散的”全新规制模式。<sup>[15]</sup>在这里，算法扮演着重要角色，“代码就是法律”。<sup>[16]</sup>司法规则的代码化和自动执行成为一种趋势，不仅消减了以往的“司法代价”，也产生了特殊的司法效果。

[13] [法] 米歇尔·福柯：《规训与惩罚》，刘北成等译，生活·读书·新知三联书店1999年版，第226页。

[14] 前引[10]，吉登斯书，第18页。

[15] [英] 詹姆斯·柯兰等：《互联网的误读》，何道宽译，中国人民大学出版社2014年版，第128页。

[16] [美] 劳伦斯·莱斯格：《代码2.0：网络空间中的法律》，李旭等译，清华大学出版社2018年版，第6页。

### 1. 自动化的程序运行

随着人工智能技术的广泛应用,自动化执法不再是一种“越轨”。<sup>[17]</sup>同样,人工智能技术也可以将司法规则转换成能够自动运行的程序代码,“按照合乎程序规定的方式将司法过程精确拆解,又借助数据分享和集中管理,令系统中的每一个动作都为最上的管理者可见”。<sup>[18]</sup>这些智能程序没有情感纠缠,一丝不苟且不知疲倦。让它们替代有人性局限的司法者,就成为新时代的一个选项。例如,上海的“206系统”建立了逮捕条件审查、证据标准指引、单一证据审查、证据链和全案证据审查、庭审示证、类案推送、办案程序合法性审查监督等自动程序。假设警察破获了一起案件,想要逮捕犯罪嫌疑人,系统会自动提示哪些关键证据缺失、哪些证据有瑕疵;检察官起诉时,系统会提示类似案件的起诉比例如何;法官将写好的判决书放入系统中,系统会自动提示判决在多大程度上同本院以及上级法院在类似案情中的判决不一样,法官若坚持该判决,系统将自动把判决推送给庭长以供讨论。<sup>[19]</sup>司法区块链技术则实现了从生成智能合约、完成实人认证并签约、合同原文及智能合约上传至司法区块链、智能合约自动运行、合约无法执行后转入多元调解流程、纳入信用奖惩联合机制、立案、审判、执行的全流程智能化。2019年12月19日,杭州互联网法院对一起该类型案件进行了宣判,这是国内首例“无人工干预、无外部因素干扰”的“智审”断案。<sup>[20]</sup>这些人工智能辅助办案系统,已初步实现了“去人化”,在一定程度上实现了司法标准的统一,且降低了成本,提高了效率。

### 2. 隐身性的司法能动

在当今信息时代,技术赋能可以提高司法效率、提供“可视正义”,同时也助长了司法能动。一是实现了空间拓展。近年来,北京互联网法院的电子诉讼平台、广州中院的“律师通”和“移动执行”手机APP、浙江“移动微法院”等纷纷上线运行。2019年广州互联网法院还在穗港澳布设了12个“E法亭”,全方位为当事人提供自助存证、自助立案、自助查询、智能送达、在线调解、在线庭审等六大诉讼服务功能。<sup>[21]</sup>这类智能平台走出了法院的固定场域,实现了在现实/虚拟双重空间中的拓展,增强了司法实现“可视正义”的能力。二是实现了权力增能。例如,通过执行网络办案平台,办案人员足不出户就能完成对被执行人财产的自动查询和冻结、扣划,直接查询被执行人在异地不动产登记详细信息和投资的具体情况;执行智能预警平台则通过汇集人口、婚姻、金融、政务、电商、信用、社交等周边数据,全方位评价当事人的身份特质、行为偏好、资产状况、信用历史,对涉案当事人进行信用画像和自动风险评估。<sup>[22]</sup>这些智能联动,使得司法权力穿透了原来的物

[17] 参见[美]瑞恩·卡洛等编:《人工智能与法律的对话》,陈吉栋等译,上海人民出版社2018年版,第239页。

[18] 芦露:《中国的法院信息化:数据、技术与管理》,载苏力主编:《法律和社会科学》第15卷第2辑,法律出版社2016年版,第46页。

[19] 见《最高法司改办何帆:中国法院正努力把人工智能引入办案系统》, [https://www.thepaper.cn/newsDetail\\_forward\\_1746283](https://www.thepaper.cn/newsDetail_forward_1746283), 2020年4月30日最后访问。

[20] 见《中国首个应用区块链智能合约技术案件宣判》, <http://www.chinanews.com/sh/2019/12-19/9038475.shtml>, 2020年4月30日最后访问。

[21] 见《是“E法亭”,不是“一法庭”!广州打通司法服务最后一公里》, <http://dy.163.com/v2/article/detail/EVII2P8D05149N9E.html>, 2020年4月30日最后访问。

[22] 见王牧、吴磊:《“智慧执行”模式的具体探索和未来发展》, <https://www.chinacourt.org/article/detail/2018/01/id/3179286.shtml>, 2020年4月29日最后访问。

理“围墙”，延伸至所有涉案领域，并在控制力上获得了增能扩张。三是实现了关联限制。在以往的司法运行中，司法机关对当事人权利的限制措施基本都是“就事论事”，并遵守相应的规则和程序。司法人工智能带来了“关联限制”效应，如采取“弹屏”电子送达方式时，法院发送到被送达人手机上的短信送达内容，不受杀毒软件、手机终端设置等方面的屏蔽拦截，无论其手机处于操作过程还是待机状态，均被即刻锁屏，查阅内容后方能解锁。这固然确保了送达的便捷性、有效性和可靠性，但如今手机所承载的功能、意义和价值已经远远超出通讯工具本身，关涉移动支付、商品交易、媒体社交、个人隐私等多方面的生活权利，一旦被锁屏，它所锁定的就是这些权利，而不仅仅是同送达相关的权利，这样就产生了关联限制。<sup>[23]</sup> 这些隐身的司法能动倾向，无疑加深了司法权的社会嵌入程度和影响。

### 3. 技术性的规则重写

司法具有鲜明的法条主义和形式主义特征，但“在一些问题上，我们没有办法事前就给定法官可以机械适用的规则，而必须由其来创制规则”。<sup>[24]</sup> 司法不仅是一个适法的过程，也是理解、重释甚至重写规则的过程。司法人工智能所带来的法律规则代码化，既有控制自由裁量、对诉讼流程进行节点控制、实现“可视正义”的司法效果，也会带来对规则进行技术性重写的机会。“代码作者越来越多地是立法者。他们决定互联网的缺省设置应当是什么；隐私是否将被保护；所允许的匿名程度；所保证的连接范围。”<sup>[25]</sup> 一旦可以把司法规则转化为代码，对代码的控制就成为在司法解释和“法官造法”之外的二次重释或重写规则的权力。这些代码的编写基本都是由司法机关和技术专家主导，很少有律师、学者和公众参与。司法者难免从有利于司法权力运行的立场出发，在数据标注、要素抽取和代码编写过程中注入自身的或者职业的价值偏好，这意味着在数字司法向度上扩大了司法权。

#### （三）司法决策建模化

在当今世界，数字经济已成为主导形态，“人类正在进入一切皆可计算的时代”。<sup>[26]</sup> 与此相应，社会生产和生活都逐渐走出了个别性、分散性、经验性的决策方式，形成了体系化、自动化的建模运营方式，通过信息采集、数据分析和设计算法来实现决策的智能化，从而“提供了一种全新的革命性工具”。<sup>[27]</sup> 这意味着，很多决策“已经从人手中转移到算法手中”。<sup>[28]</sup> 智慧司法的一个重要方向，就是司法决策的建模化。

首先，生成法官智慧的聚合效应。理性选择理论上的“聚合效应”或者“聚合的奇迹”，意指在规模群体的民主决策过程中，个体计算或预测的误差会彼此抵消。“在绝大多数情况下，平均意味着平庸。但在决策时，却常常效果非凡”，<sup>[29]</sup> 很容易形成意想不到的、

[23] 见祝文明：《北京互联网法院推出弹屏短信送达》，[http://www.cipnews.com.cn/Index\\_NewsContent.aspx?newsId=1111737](http://www.cipnews.com.cn/Index_NewsContent.aspx?newsId=1111737)，2020年4月29日最后访问。

[24] [美] 李·爱泼斯坦等：《法官如何行为：理性选择的理论和经验研究》，黄韬译，法律出版社2017年版，第25页。

[25] 前引[16]，莱斯格书，第87页。

[26] 前引[9]，徐格等书，第323页。

[27] [英] 乔治·扎卡达基斯：《人类的终极命运：从旧石器时代到人工智能的未来》，陈朝译，中信出版社2017年版，第296页。

[28] [美] 克里斯托弗·斯坦纳：《算法帝国》，李筱莹译，人民邮电出版社2014年版，第197页。

[29] [美] 布赖恩·卡普兰：《理性选民的神话：为何民主制度选择不良政策》，刘艳红译，上海人民出版社2010年版，第7页。

最优的多数决判断。然而，人们很快又发现了它的局限。由于聚合过程中充满了“系统性错误”，不可通约的偏见“污染”了聚合效应，简单过半数投票在加总个人偏好并排序后，“会因出现严格社会偏好的循环，导致无法从一些有限的可行社会状态集中确定一个最优的社会状态”。<sup>[30]</sup>而司法人工智能中的决策建模化，能够清除这种可能的“系统性错误”，进而生成法官智慧的聚合效应。

具言之，虽然有公开的明确规则、统一的司法政策和遵循先例原则，但“法律的生命不是逻辑，而是经验”。<sup>[31]</sup>面对日常生活中种类繁多、情节复杂的大量案件，这些规则、政策和原则难以直接被无缝套用，自由裁量不可避免。即便是最智慧、最敬业、最高尚的法官，也只能作出分散性、个体性的裁量，无法克服个体决策的有限性。如何约束自由裁量权以提升裁判水平、促进公平正义就成为一个难题。司法人工智能通过数据标注识别、案件要素抽取、知识图谱构建来进行算法建模，这个“机器人法官”拥有的不再是法官个体或者合议庭法官的有限能力和个体经验，而是来自海量类似案件信息及裁判结果的数据库运算，它所作的每一份判决，都是一项数据庞大的实证研究和综合衡量。例如，在上海“206系统”运用深度学习算法构建的智能量刑模型中，仅盗窃罪一个量刑模型的学习样本，就囊括了全国三十万个盗窃案件样本的全部案情要素及量刑情节要素。<sup>[32]</sup>这样，就形成了一种空前的聚合效应。由于这种建模算法并不是基于一个个法官的即时投票，而是基于业已形成的个案化的既定裁判，并且这些裁判都是立足相同职业背景和专业知识的理性判断，不易受到普遍的、不可通约的偏见“污染”。此时，才真正有可能创造出“聚合的奇迹”，这相当于整合、提炼了所有已决案件法官的集体智慧和最佳方案。

其次，加持数理逻辑和计算理性。现代性法律的一个突出特征是形式理性。它体现着“在技术上可能的计算和由它真正应用的计算的程度”，<sup>[33]</sup>推动了法律规则的权利义务设计和利益平衡，使人们能够预测自己和他人的行为效果，合理规划自己的行为。而法官在案件裁判过程中，则存在着“法官造法”的构建理性和自由裁量的经验理性，他们力图寻求逻辑一致、唯一正确的司法结果，进而维持法律的形式理性。

从生命科学来看，人脑思维本身也是一种问题求解的计算过程。法律或司法上的各种理性形式就是各种生化算法。如今信息技术能够写出越来越复杂的电子算法，二者可能会适用相同的数学定律，这就“打破了动物和机器之间的隔阂，并期待电子算法终有一天能够解开甚至超越生化算法”。<sup>[34]</sup>其实，在20世纪70年代，人们就通过“如果A和B，那么C”的方式对三段论加以描述，让机器法律推理变成可操作的现实。“将法学问题形式化为可计算问题”，<sup>[35]</sup>然后选择合适的计算方法进行建模编程，就成为一种智慧司法的潮流，英国“在线法院”法官还基此提出了“算法裁判”的倡议。<sup>[36]</sup>在我国，司法人工智能主要

[30] [美] 埃里克·马斯金等：《选择的悖论：阿罗不可能定理与社会选择真相》，黄永译，中信出版集团2016年版，导言，第XV页。

[31] [美] 小奥利弗·温德尔·霍姆斯：《普通法》，冉昊等译，中国政法大学出版社2006年版，第1页。

[32] 参见前引[11]，崔亚东书，第168页。

[33] [德] 马克斯·韦伯：《经济与社会》上卷，林荣远译，商务印书馆1997年版，第106页。

[34] [以色列] 尤瓦尔·赫拉利：《未来简史》，林俊宏译，中信出版社2017年版，第333页。

[35] 邓矜婷、张建悦：《计算法学：作为一种新的法学研究方法》，《法学》2019年第4期，第104页以下。

[36] 参见曹建峰：《“人工智能+法律”十大趋势》，《机器人产业》2017年第5期，第86页以下。

通过机器学习来构建司法知识图谱和算法模型，可以在线一键启动从开庭到裁判执行的所有流程，也即“采用大数据和诉讼建模将管辖法院、诉讼请求、赔偿数额的计算、法律依据的引用等诉讼事项进行全面模块化，当事人选项即可完成起诉、应诉等过程”。<sup>[37]</sup> 在这种情况下，智能系统的规范性含义会变得更加清晰，“它们平均会比人类专家提供更好的结果”。<sup>[38]</sup> 这就在法律表达和运行上加持了数理逻辑和计算理性，把知识系统转化成了数字系统，把案件审理过程转化成了案件运算过程，从而提供更客观、更科学、更准确的纠纷解决方案。

再次，实现同案同标、同案同判。近代以来，司法被赋予独立品格和中立立场，但面临的一个根本困境是：如果法官恪守条文，机械地司法，可能无法回应社会需求、实现公平正义；如果法官们偏离法律过于频繁，“违反了选择最优化，则法治便无法存活”。<sup>[39]</sup> 更为复杂的是，面对类似或相同案件，每个司法者的判断都会有所不同，他们可以轻而易举地，即使是无意识地，以发现法律规则本意的名义，在裁判中推行他们自身的偏好。<sup>[40]</sup> 要摆脱上述困境，关键在于克服法官自由裁量中的先入之见、价值偏好，以及心情、性格和精力等主观因素的影响。尽管真正意义上的同案或同判可能都不存在，人们却一直渴望司法统一、同案同标、同案同判，只不过受诸多主客观因素或条件限制，这一期待常常落空。

司法人工智能为人们化解这一难题带来了巨大希望。按部就班的程序没有情感和意识，更没有先入之见、价值偏好等因素的干扰。从理论上讲，“算法可以详细地描述应用于每个小的决定（进而形成最终的重要决定）之中的那些具体规则”。<sup>[41]</sup> 它们以标准一致的要素抽取、相同的算法建模以及流水化的操作流程，对相同的信息输入给出相同的算法输出。在这种智能系统自动运行的司法决策条件下，同案同标、同案同判将有机会走进现实生活。

#### （四）司法服务管理智慧化

随着网络化、数字化、智能化技术在司法领域的广泛应用，“让当事人少跑路、让数据多跑路”，从“最多跑一次”转向“一次都不用跑”，不再是一句口号。2019年，全国97.8%的法院支持网上立案，其中高级法院的支持率达100%；全国法院支持网上证据交换的达66.9%，支持网上开庭的达58.2%；全国高级法院公开终本案件信息的比例达100%，中级法院和基层法院也分别达98.8%和96.6%。当事人通过中国执行信息公开网共联系法官73203次，法官及时回复率为85.2%。<sup>[42]</sup> 这些都大大提升了司法服务的智慧化、便捷化，经由技术赋能实现了“服务群众诉讼全方位、全天候、零距离、无障碍”的目标。

司法管理和司法监督也实现了智慧化、透明化。例如，上海“206”系统采用了审判管

[37] 邓恒：《人工智能技术运用与司法创新》，《人民法院报》2017年12月14日第2版。

[38] 前引[17]，卡洛等编书，第120页。

[39] [美]杰弗里·布兰德：《法治的界限：越法裁判的伦理》，娄曲亢译，中国人民大学出版社2016年版，第313页。

[40] 参见[美]欧文·费斯：《如法所能》，师帅译，中国政法大学出版社2008年版，第13页。

[41] [美]卢克·多梅尔：《算法时代：新经济的新引擎》，胡小锐等译，中信出版社2016年版，第139页。

[42] 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：《2019年中国法院信息化发展与2020年展望》，载陈甦、田禾主编：《法治蓝皮书：中国法院信息化发展报告 No.4（2020）》，社会科学文献出版社2020年版，第2页以下。

理数字化智能提示、审判执行流程可视化智能跟踪管理、审判质效目标化动态管理、法官业绩数字化智能评价等技术,并进一步运用大数据分析、视频图像识别、语义分析等技术,实现了办案过程全程可视、全程留痕、全程监督,确保司法权的阳光运行。借助这些数据分析和智能技术,的确更容易发现恶意诉讼,以及同私下交易有关的“关系案”“人情案”“金钱案”,庭审中的迟到、退庭、衣冠不整等现象也都会透明可视。

总之,司法人工智能使司法运行产生了前所未有的深刻变革。在司法效果上,实现了司法权在物理空间中的去中心化和扁平化,以及在电子空间中的能动延展和扩张,并有可能出现数字司法界面下的“再中心化”。在司法逻辑上,形成了知识系统与数字系统、法律理性和计算理性的互动构建。建立在思辨的逻辑推理基础上的“公理思维”,开始转向立足计算和相关性分析的“数据思维”。在司法机制上,智能辅助办案系统的平台化,使得具有分工制约关系的公检法司开始高效一体化运行。在司法裁决上,出现了正义的程式化、产品化供给方式,算法也从“提炼自这个世界,来源于这个世界”,转向“开始塑造这个世界”。<sup>[43]</sup>这些无疑都使司法获得了巨大的技术赋能,推动着“接近正义”向“可视正义”转型。

### 三、司法人工智能的内在限度

人工智能技术融入司法是数字时代的必然趋势,其发展前景无需质疑,但“要想让算法更好地服务于人类,我们必须承认算法不是全能的”。<sup>[44]</sup>智慧司法建设并不是一项司法革新的完美规划,它对司法机制及其运行的重塑效应也是有限的。

#### (一) 客观性上的悖论

我国的智慧司法建设,旨在实现“全面覆盖、移动互联、跨界融合、深度应用、透明便民、安全可控”的转型升级,力图以算法决策规避人脑的“决策噪音”,让司法活动更为阳光,司法公正更为直观,最终提升司法结果的客观性、一致性、公正性和效率性。然而,从全球范围的实践来看,算法决策和司法大数据的客观性并不像人们想象的那样易得,其也存在被主观因素侵入的风险。

##### 1. 算法决策的复杂性

首先,算法决策是数字时代的新型权力。人们曾把算法视为一种自由言论或商业秘密,但这些都不是算法的本质所在。人工智能在商业交易、公共政策的制定、司法过程、交通出行等日常生活中深度应用的场景越来越多,它们不仅仅进行预测和推荐,很多时候也在作定性判断,如人脸识别、资格审查、屏蔽封号、自动执法等。“这些自动化系统已经由简单的行政管理工具变成了主要的‘决策者’”,<sup>[45]</sup>形成了一种替代人脑决策的裁断权力,并在资源配置和行为控制上对相对人产生重要影响。就智能辅助办案系统而言,它所进行的证据标准和规则指引、逮捕条件审查、证据链和全案证据审查判断、社会危险性评估、言

[43] 前引[41],多梅尔书,第214页。

[44] [美]凯西·奥尼尔:《算法霸权——数学杀伤性武器的威胁与不公》,马青玲译,中信出版集团2018年版,第243页。

[45] 前引[41],多梅尔书,第137页。

词证据审查、减刑假释案件办理、裁判偏离度提示、办案程序监督等等，也都是在行使算法裁断权力。这种算法决策，并非简单的计算公式和无意识的程序运行，而是从一开始就被“教导”了法律知识与逻辑，输入了主观的定性判断基准。尽管它需经司法裁判者最终确认，才能产生实际的司法效力，但仍不可否认其权力属性。

其次，算法权力也会受价值偏好影响。人们曾对算法决策抱有技术中立和进步福利的浪漫幻想，认为它在准确性、一致性和效率上远优于人类，而且不受情感和私欲干扰。“对于计算机来说，除非程序设计者把偏见因素引入进来，否则它不会存在这个问题，而且我们可以通过查看程序的内容来判断其是否存在偏见因素”，<sup>[46]</sup>但问题也恰恰出现在这里。设计者“可以建造、构筑、编制网络空间，使之保护我们最基本的价值理念，同样地，我们也可以建造、构筑、编制网络空间，使这些价值理念丧失殆尽”。代码是被制造而非发掘出来的，并且“仅由我们所制造”。<sup>[47]</sup>基于相同的数据信息、技术条件、求解方案，分别由两组各自封闭的程序员来设计算法，最后的输出结果很可能不一样。甚至在不同的使用场合，当人们使用完全相同的运算步骤，想要得出相同结论，都常常不太可能。<sup>[48]</sup>这也说明了设计者主观性的重要影响。尽管问题能够被发现，但很难通过程序审查进行纠偏。2018年初美国纽约市出台“自动化决策特别工作组法”，试图对政府使用的各种算法进行监管，最后不了了之。<sup>[49]</sup>同样，在我国的人工智能司法应用过程中，司法机关各自或者联合研发的智能系统，都不同程度地反映着部门的职权立场、行动逻辑和地域特点。例如，上海“206系统”工程庞大复杂，只能采取全国法院的“众筹”模式来完成证据知识图谱的绘制。这些法律适用规则的整理，其实就是对刑法知识的一次精加工，“编辑者的逻辑编排、要旨提炼、观点选择，体现了个人的价值取向、学术判断、政策立场”。<sup>[50]</sup>这些知识图谱经由程序员的代码编写和算法建模，自然又会或多或少地渗入一些人为因素。尽管代码的执行是无偏差的，但“偏差”会被编码到系统中。<sup>[51]</sup>

再次，算法背后隐藏着算法黑箱与“算法霸权”。自人类社会开启了“算法为王”的智慧发展模式，建模算法开始“大胆地从混沌之中创建秩序”。<sup>[52]</sup>然而，它们在提供技术进步福利的同时，也带来了算法黑箱问题。“用户们无法看清其中的规则，无法提出不同意见，也不能参与决策过程，只能接受最终的结果。”<sup>[53]</sup>这就形成了一个算法决策的闭合回路，除非自我更新和改善，否则它将持续运行，不受外界因素的干扰和影响。一旦算法决策出现偏见、不公乃至有意的“暗算”，就会诱发严重的社会危机。有研究认为，“程序员和设计者可能会试图以公开或隐蔽方式嵌入一种逻辑，以使算法偏向期望的目标”，如为特

[46] 前引〔24〕，爱泼斯坦等书，第425页。

[47] 前引〔16〕，莱斯格书，第6页。

[48] 参见〔德〕罗纳德·巴赫曼等：《大数据时代下半场：数据治理、驱动与变现》，刘志则等译，北京联合出版公司2017年版，第263页。

[49] 参见张凌寒、李荣：《美国纽约算法监管遇挫启示录》，《法治周末》2020年1月16日第11版。

[50] 何帆：《人工智能时代，为何还要打造这样一本刑法工具书》，<https://new.qq.com/omn/20190508/20190508A0708Q.html>，2020年4月6日最后访问。

[51] 参见前引〔17〕，卡洛等编书，第283页。

[52] 〔美〕Robert C. Martin：《代码整洁之道：程序员的职业素养》，余晟等译，人民邮电出版社2016年版，第18页。

[53] 前引〔41〕，多梅尔书，第139页以下。

权阶层布设“后门”。<sup>[54]</sup> 尽管算法权力以建模计算之必然输出结果的形式, 获得了技术中立的外在表现, 但它不予公开、不接受质询、不提供解释、不进行救济, 难免会有演化为“算法霸权”的风险。<sup>[55]</sup> 于是, 算法解释权和算法治理随之成为一种时代诉求。<sup>[56]</sup> 我国的智慧司法建设也应着力消除算法黑箱和算法霸权的隐患。未来在程序设计环节实现立法者、专家学者、律师和社会人士的开放性参与, 打造可公开、可解释、可救济的智能系统, 对于防范算法黑箱和算法霸权之风险, 大有必要。

## 2. 司法大数据中的主观性因素

从19世纪开始, 一些科学家就期望利用客观、可靠的统计数据来“构建完美社会”,<sup>[57]</sup> 司法大数据的挖掘利用也旨在打造更理想的司法效果。然而, 这一过程依然可能受到主观因素的干扰。

首先, 数据采集的“全样本”是一种主观界定。算法和算力仰赖于“数据喂养”, 数据量越大、越全、越真, 算法决策就越客观精准, 人工智能也才越“聪明”。司法是一项非常复杂的定分止争机制, 涉及人财物、时间地点、主客观状态、行为方式、涉案工具等等, 仅一个案件就会产生大量的司法数据。如果想就某一类型或某一类案由的案件进行算法建模, 就需要采集海量的数据。若要做到“全样本”则难上加难, 仅在什么时间、空间、范围上来界定这个“全样本”, 就是一个很大的问题。从上海“206系统”的研发过程来看, 命案被划分为现场目击型、现场留痕型、认罪供述得到印证型和拒不认罪型四大类, 这四类命案“证据标准指引”的建模算法, 主要是基于上海市2012年至2016年间审结的591起命案数据。71个常涉刑事案件罪名的“证据标准指引”, 则是基于近三年全国102个常涉罪名刑事案件的数据制定的。可以发现, 这些数据样本都是三五年内的。同样, 上海民事、行政案件智能辅助办案系统从467个民商事案由、61个行政案由中首选了六大类8个案由, 虽然标注电子卷宗达5800余份, 标注点数量达12万个, 但上海市2016年道路交通事故纠纷一个案由收案量就达51312件。<sup>[58]</sup> 可见, 研发智能辅助办案系统所凭借的, 暂时还只能是有限时间段内、有限范围内的“样本”。只要对这个“样本”的时间段和范围稍作调整, 它所汇集的司法数据情况可能就大相径庭。数量变化还是小事情, 对法律规则、司法解释、司法政策、政治事件、社会环境等方面变革及其影响的权重计量, 则更为复杂。真正意义上的“全样本”收集很难做到, 而有限时间段和范围的“全样本”选取, 很大程度上是带有主观性的。

其次, 数据标注是一种主观选择。数据标注是司法知识图谱构建的重要基础, 上海“206系统”的研发采取了人工标注和自动标注两种形式, 人工的数据清洗和标注自不待言, 机器的自动标注也需要经过程序设计人员的确认。同样, 图像识别、手写体识别、插图签章、涂抹块检测、插入检测等的准确率虽然可达92%—98%, 但它仍需人工审查校正, 这些无疑也都是主观选择的过程。

[54] 前引[17], 卡洛等编书, 第297页。

[55] 参见前引[44], 奥尼尔书, 前言, 第V页。

[56] 参见张凌寒:《商业自动化决策的算法解释权研究》,《法律科学》2018年第3期,第65页以下;张欣:《算法解释权与算法治理路径研究》,《中外法学》2019年第6期,第1425页以下。

[57] 参见前引[41], 多梅尔书, 第206页。

[58] 参见前引[11], 崔亚东书, 第128页以下。

再次，司法数据的阐释理解是一种主观判断。标定数据对机器学习和建模算法的效果具有很大影响。由于中文语词具有多义性、模糊性和地方性，对于相同或者相似的文本知识和数据，不同的标注人员会形成不同的阐释理解，导致数据标注结果可能因人而异。虽然可以通过构建多层次数据标注体系来消除语义分歧和增进标注的可解释性，但仍有一定的主观作用空间。主观性和创造性在数据分析和建模算法中发挥着重要作用，“同样的数据在不同的环境内容中可以有截然不同的意义，这些意义并非数据所固有，而是人们在特定环境中分析数据并将意义赋予了数据”。<sup>[59]</sup>这也正是“投入多少人工，就有多少智能”的原因所在。

由上可见，在科学技术发展到一定程度或治理能力足以防范相关风险之前，算法决策的客观性与中立性都只能是有限的。也许有人争辩说，人脑决策也有“人脑黑箱”，价值偏好更明显，<sup>[60]</sup>相比而言，“即使最糟糕的数学模型应该也没那么糟糕”。<sup>[61]</sup>但是，人脑决策基本都是个体化的、局部的、发散的，其反馈回路是开放的，能够反思和纠偏。而算法决策一旦出现问题，就会形成自动化、体系化、机制化的普遍侵害后果。更严重的是，算法决策往往以“科学”面目呈现，受到影响的相对人对其既不知情又不能质疑。一旦算法决策变为一种被灌输的、给定的“客观”，造成“假为真时真亦假”的反向替代，司法也就异化成为它的反面了。这是二者最深层的区别，也是我们必须警醒的地方。

## （二）正义判断上的困境

司法人工智能让人们看到了“可视正义”和“数字正义”的希望，<sup>[62]</sup>但进一步分析就会发现，“可视正义”也好，“数字正义”也罢，虽然近在咫尺，但也不会来得那么容易。

其一，难以对正义进行建模计算。正义是人类社会自古以来的向善追求和崇高价值，也是法律价值的核心所在。追求正义的理想和社会不公的现实之间的矛盾，一直是困扰人类秩序的重大难题。尽管人工智能技术可以大大提升司法实现正义的功能，但很多复杂疑难案件的正义性判断，仍然难以通过建模计算获得。

首先，正义价值是普遍的，却没有统一标准。古今中外，人人都渴望正义，但“正义本身包含着各种相互冲突的价值”，它“并不只是以一种声音说话”。<sup>[63]</sup>前些年我国发生的“刻章救妻案”，<sup>[64]</sup>就存在复杂的情理法的纠缠，“入狱养老案”也同样让人反思颇多。<sup>[65]</sup>正义呈现出一种无形的、流动的、多元的主观判断状态，“具有着一张普洛透斯似的脸”。<sup>[66]</sup>对待正义问题，只能通过谋求最大公约数和重叠共识方式来获得认同。我们很难针对正义

[59] 前引〔48〕，巴赫曼等书，第205页。

[60] 参见前引〔40〕，费斯书，第93页。

[61] 前引〔44〕，奥尼尔书，第238页。

[62] 参见前引〔3〕，凯什等书，第263页。

[63] [英]约翰·格雷：《自由主义的两张面孔》，顾爱彬等译，江苏人民出版社2002年版，第7页以下。

[64] 为救尿毒症妻子，廖丹私刻医院收费章，四年间涉嫌骗取医院17万余元。2012年12月7日，北京市东城区人民法院一审以诈骗罪判处廖丹有期徒刑3年，缓刑4年。见邓海建：《“刻章救妻”案中情与法的谦抑》，<http://opinion.cntv.cn/sikegongzhang/>，2020年4月11日最后访问。

[65] 69岁老汉生活无助，为“入狱养老”，持刀在北京站广场抢劫。2008年12月被北京铁路运输法院判处2年有期徒刑。见李代祥：《老汉入狱为养老，消除“无奈犯罪”根源更迫切》，<https://www.chinacourt.org/article/detail/2008/12/id/337941.shtml>，2020年4月11日最后访问。

[66] [美]E.博登海默：《法理学——法哲学及其方法》，邓正来等译，华夏出版社1987年版，第238页以下。

进行共识性的要素抽取和建立通用的知识图谱，“其价值难以用数字体现”，<sup>[67]</sup>也就很难进行建模计算。同时，“机器人没有道德观。它们只是一种物体，被灌输了程序员为达到特定目的而编写的代码”。<sup>[68]</sup>它所运行的数学公式，也难以为复杂的正义判断提供可靠的答案。

其次，正义没有价值位阶，裁判需要权宜平衡。正义具有丰富的价值内涵和理论体系，却没有一个明确的位阶图表，“它意味着在各种善之间不存在一种完备的字典式排序；它还意味着不存在‘社会制度的首要美德’，而只是一系列公共的善和美德，由于环境的不同而具有相对的重要性”。<sup>[69]</sup>关键问题并不是确定哪一个价值、哪一种善更重要，而在于“维护这样一种制度：对于过一种良善生活是重要的所有价值的要求都能在这种制度中得到考虑和权衡”。<sup>[70]</sup>因此，比例原则、法益衡量等一直在司法裁判中占据重要位置。在社会公众心中，“法官应当被视为两个世界永恒的杠杆：应然的世界和实然的世界，公共价值的世界和主观偏好的世界，宪法的世界和政治的世界。法官在其中之一寻求合法性，但必须在另一个中认识自身”。<sup>[71]</sup>司法中的算法决策，必须能够面对和处理这样的权衡情境。至少，“当任务是构造一个正确的或公正的法律裁决时，或者是对这样的一个裁决进行评论时，相关的各个当事人的利益和观点必须作为模型的一部分”。<sup>[72]</sup>这就很有可能超出了人工智能的建模计算能力。当然，科技中的想象力可以让人受到鼓舞，人们期待能在智能系统中嵌入一种“道德算法”，或者具有道德推理和道德讨论能力的计算系统，以解决伦理判断难题，比如“如果一个机器人开展搜救工作，那么它应该先疏散/救援谁？或者，如果提供社会陪伴服务，那它什么时候该和用户撒谎”。<sup>[73]</sup>然而，日常伦理判断毕竟不同于司法裁判。道德算法或许能够应对日常伦理判断，若要处理情节复杂的重大疑难案件，道德算法还是难以胜任。从目前看，“法律推理的所有关系模型都不能用于恰当地建立司法自由裁量权的模型”，<sup>[74]</sup>力图解决一切、无所不能的“终极算法”也还不存在。司法裁判中法官的权宜平衡一时还难以用算法来替代。

再次，实现正义依靠的是实质判断，而不是体现相关性的概率计算。随着人工智能技术快速发展，司法领域中算法决策与人脑决策的对弈测试也逐渐增多。美国曾研究测试，人工智能在584个案件中有79%的案件审理结果与欧洲人权法院的审判专家组判决结果相同。英国“机器律师”与伦敦100名律师在保险合同案件的法律判断比赛中，“机器律师”以86.6%的准确率领先于律师的66.3%。<sup>[75]</sup>这些事例似乎表明，算法决策比人脑决策更优越。但实际情况是，“计算机并不是‘思考’，而是依据逻辑算法处理数据”。<sup>[76]</sup>它主要是基于数据进行相关性分析，但这些“相关性关系并非决定性的，而是随机的，属于概率

[67] 前引[44]，奥尼尔书，第242页。

[68] [美]约翰·C.黑文斯：《失控的未来》，全琳译，中信出版社2017年版，第22页。

[69] [美]威廉·A.盖尔斯敦：《自由多元主义》，佟德志等译，江苏人民出版社2005年版，第6页。

[70] [美]约翰·凯克斯：《反对自由主义》，应奇译，江苏人民出版社2003年版，第233页。

[71] 前引[40]，费斯书，第60页。

[72] [德]托马斯·F.戈登：《诉答博弈——程序性公正的人工智能模型》，周志荣译，中国政法大学出版社2018年版，第8页。

[73] [英]玛格丽特·博登：《AI：人工智能的本质与未来》，孙诗惠译，中国人民大学出版社2017年版，第195页。

[74] 前引[72]，戈登书，第8页。

[75] 参见前引[11]，崔亚东书，第35页以下。

[76] 前引[27]，扎卡达基斯书，第292页。

逻辑”。〔77〕此时，如何用数量、概率和逻辑来表达和定义善、恶、权利、义务、公正等伦理范畴，通过这种代码转换使人的伦理变成程序化的机器伦理，〔78〕就显得十分关键。“我们必须将意图写入法律，同时写入控制机器的软件。”〔79〕这对于简单案件而言，也许并不难做到，但对于复杂的疑难案件来说，仅有这种数量、概率和相关性逻辑还远远不够，它们更多需要因果性分析。这种因果性分析并非单纯的形式逻辑演绎，而是一种实质性判断。诚如霍姆斯所言，“一个时代为人们感受到的需求、主流道德和政治理论、对公共政策的直觉——无论是公开宣布的还是下意识的，甚至是法官与其同胞们共有的偏见，在决定赖以治理人们的规则方面的作用都比三段论推理大得多”。〔80〕正义的司法裁判需要通过综合考量各种因素的实质性判断来实现，而很难依靠计算概率实现，即便算法可以进行形式逻辑推演，也无法作出这种实质性判断。

其二，难以穷尽“正义”的样本。拥有了司法正义的“全样本”，才可能实现精准的算法正义。理论上讲，这个司法知识图谱应该是法律法规、司法文件、法院判例、证据规则和案件事实的动态集合，其数据采集是可以做到的。但实际上，大量的司法活动难以被完全数据化，且各机关或机构的数据处理标准常常不一致，数据交换也有障碍。即使在单个的刑事罪名背后，也隐藏着千百种“适用场景”，对应着各类成文或不成文规则，既有法律适用规则、量刑操作规则，也有证据审核规则、程序把关规则等等。如果没有专业人员进行提炼、分类、整合和标准化处理，进而将之转化为算法嵌入系统，那就只能作“打包”处理。即使不“打包”而是进行详尽的规范化处理，也无法涵盖司法实践中的各类复杂情况。例如，对“入户抢劫”的理解，就需结合数个个案情形，依据相似性判断、有效性判断、相关性判断和技术性判断来认定。这表明，要对司法正义的各类情形和所有样本作出穷尽式的提炼与归纳，操作起来十分困难。这还没有把不同时间段的法律修订、法律解释改变、司法政策变化、潜在冤错案、案例内容矛盾、司法的地方性差别等因素考虑进去，否则将会更加复杂，也更难穷尽案件的正义样本。

其三，正义难以零和博弈。从根本上说，“算法就是把输入转换成输出的计算步骤的一个序列”。〔81〕面对海量的司法大数据、复杂的法律规则体系和司法运行机制，算法的开发设计者需要把知识系统转化为数字系统，把法律逻辑转化为数理逻辑，进而把复杂的司法活动简化为一种“无须满足任何更多的条件即可生成‘是’与‘否’的二元选项”的代码执行和机器运算。〔82〕这固然提高了司法效率和裁判一致性，但不容否认的是，“法律世界包罗万象、复杂混沌，常常没有明确的对错之分”，〔83〕尤其是司法正义，更不可能简化为一个是与否的零和博弈。算法决策“是”或“否”的对立判断，不免存在风险。“一旦它们

〔77〕 前引〔48〕，巴赫曼等书，第199页。

〔78〕 参见段伟文：《人工智能的道德代码与伦理嵌入》，《光明日报》2017年9月4日第15版。

〔79〕 [德] 克里斯多夫·库克里克：《微粒社会——数字化时代的社会模式》，黄昆等译，中信出版社2017年版，第136页以下。

〔80〕 前引〔31〕，霍姆斯书，第1页。

〔81〕 [美] 托马斯·科尔曼等：《算法导论》，殷建平等译，机械工业出版社2013年版，第3页。

〔82〕 前引〔41〕，多梅尔书，第220页。

〔83〕 陈尔彦：《法律如何应对人工智能挑战》，《经济参考报》2017年12月12日第8版。

做出了错误的决定，其造成的后果就很有可能是骇人听闻的。”<sup>〔84〕</sup>此时，司法正义就会面临严峻的挑战。

其四，正义难以忽略人性。从古到今，司法活动都不是一个冰冷的、机械的规则适用过程，而是带有鲜明的人文关怀。人们追求司法正义，无疑也是为了塑造更加公平和合理的社会秩序。尽管法官有人性的弱点，自由裁量权有专断腐败的可能，但他们的经验正义理念和司法主动性也确实抵消了“不断增多的普遍性规则可能产生的僵化行为”，<sup>〔85〕</sup>维护着司法公正的公信力。相比之下，司法人工智能固然有助于作出更为高效一致的司法裁判，但它们不会计算情感，不能与人进行心灵互动，自然也不可能有人文关怀，难以根据具体场景灵活地、创造性地维护正义价值。而越来越多的机器执法和裁断，越来越少的人际互动和交流，有把公民变成“驯服的身体”的危险。因此，在通过算法决策实现司法正义的过程中，需要人类“为系统提供必要的情感和共鸣元素”。<sup>〔86〕</sup>“审判中不可减少的人性，恰恰在于人类对事实和矛盾的法律要素审查所产生的不可预测性，这种不可预测性或多或少取决于不同情况。”<sup>〔87〕</sup>目前各国大多都将算法决策定位为辅助系统，办案人员必须处于主导地位，以防止算法决策带来人文流失。可见，司法人工智能只能带来有条件的“可视正义”和“数字正义”，即使法律人工智能已广泛投入运用，在追求司法正义的过程中，“法律人的‘情怀’和‘匠心’”仍然无法被复制和替代。<sup>〔88〕</sup>

### （三）政治因素的遮蔽

现代法治的核心特征之一是司法中立，然而，司法毕竟是一种制度体系中的构架，它不可能做到隔绝政治。“法院在某些情况下不得不处理社会政策问题，这并不是因为人们特别希望他们这么做，而是因为通常情况下法院没有别的选择。”<sup>〔89〕</sup>例如，美国联邦最高法院可以通过重新定义某些规则条款和标准，“来满足大法官不断变化的政策偏好，从而在集权与分权、保守与自由之间做出最有利于大法官的选择”，<sup>〔90〕</sup>但“布什诉戈尔案”之类的司法裁决仍不是一种单纯的司法行为，而是带有明显的政治考量。<sup>〔91〕</sup>基于我国的制度属性和历史传统，政治性因素对司法的影响也比较明显。比如，“民愤极大”就曾是一种从政治上、道义上来考虑的从重处罚情节，“领导批示”也会对案件的定性和裁判具有重要“指导”作用。

这些影响因素很难进入算法决策的考量范围。首先，政策性、政治性影响因素往往是无形的、不确定的，如“坚持法律效果、政治效果、社会效果相统一”的综合考量，就潜在在于办案人员某种未能言明的“自由心证”之中。其次，社会舆情、领导批示和联合专案

〔84〕 前引〔41〕，多梅尔书，第136页。

〔85〕 前引〔40〕，费斯书，第96页。

〔86〕 前引〔17〕，卡洛等编书，第264页以下。

〔87〕 [法]伊莎贝尔·贝林：《法院判决真可以用数据预测吗》，李蒙译，<https://www.chainnews.com/articles/678354421816.htm>，2020年4月6日最后访问。

〔88〕 参见前引〔50〕，何帆文。

〔89〕 前引〔24〕，爱泼斯坦等书，第22页。

〔90〕 [美]杰弗瑞·A.西格尔等：《正义背后的意识形态：最高法院与态度模型》，刘哲玮译，北京大学出版社2012年版，第36页。

〔91〕 参见韩显阳：《美国最高法院大法官席位争夺战开打》，《光明日报》2016年2月17日第12版。

组等一些政治性因素，并不能全部反映到卷宗中，很多都反映在办案人员的政治立场和方向思路上，很难做到全程留痕。再次，办案人员往往会利用法律技术和规则漏洞，采取表面上遵从司法程序、符合法律规定、体现法律逻辑的形式，将政治性因素予以融入吸收并“合法化”，进而将其“遮蔽”在司法过程和司法裁判中。对这些无影无形却又有重要影响的因素，很难进行数据采集、样本分类、要素抽取和知识图谱构建，也难以将其纳入建模计算。

#### （四）精准性上的障碍

人工智能之所以被广泛应用于生活领域，原因之一就是它具备远超人脑的精准计算和超强预测能力。但在实际运行中，算法决策未必能完全做到精准，甚至还有一些风险需要加以防范。一是算法错误。人工智能虽然都有“先发制人”和“完美预防”的预设，但无缺点的自动化系统并不存在，<sup>[92]</sup>需要容忍不可避免的系统错误和故障。经验表明，“由于程序员偷懒或者疏忽，不少被编码的法规与书面法规相比，确实存在精细度不够的问题”，而“错误的编码导致法律条文被严重扭曲，其效果也发生了改变”。<sup>[93]</sup>二是算法自主。随着人工智能技术的发展，人工智能在分析、预测和判断上的某些能力已经超过了人类。有时还会出现算法超出设计预期的运算方式和输出结果，令设计者难以理解的情况，即算法自主或者算法异化现象。三是算法不能。其实从图灵开始，人们就已经意识到，虽然在很多情况下算法执行的任务是人类无法完成的，但算法并不能解决所有问题，<sup>[94]</sup>如不少刑事案件中犯罪者的动机和目的。四是算法差异。在人工智能的实际应用中，相同的任务或应用可以采用不同的算法来完成，但这些不同算法的正确性、容错性和效率性存在优劣之别。从审级上说，从基层人民法院到最高人民法院，究竟是使用一套算法还是四套算法？如果使用一套算法，就会导致四个审级的智能裁判结果完全一致，不同审级之间的监督功能就会丧失；如果使用四套算法，那又如何保证它们的计算品质相同？从地域上说，全国如果使用一套算法，要如何解决经济、社会和文化上的地域差异问题？如果使用多套算法，如何保证全国的司法统一性？可见，尽管算法决策是基于客观的数据分析和理性计算作出的，但面对重大疑难案件和大国的司法运行，有很多复杂因素难以用建模算法来回应。即使可以进行量化计算，也可能达不到人们所期待的精准程度。“最难诉诸自动化的任务，就是那些要求灵活性、判断力和常识技能等人只能通过默契来理解的任务。”<sup>[95]</sup>

## 结 语

本文对司法人工智能的正向展望与反向省思，并非是“以己之矛、攻己之盾”的逻辑游戏，而是旨在深入分析和研究它的可能空间与内在限度。对司法人工智能的进步与前景，我们当然抱有积极乐观的态度，但越是这样就越要对其存在的问题和自身限度予以高度重

[92] 参见前引〔17〕，卡洛等编书，第257页。

[93] 前引〔41〕，多梅尔书，第140页以下。

[94] 参见〔澳〕托比·沃尔什：《人工智能会取代人类吗？——智能时代的人类未来》，阎佳译，北京联合出版公司2018年版，第12页。

[95] 同上书，第84页。

视, 从而进行合理定位和科学赋能。发挥人工智能对司法的优化重塑效应, 与对人工智能的风险防范应同步进行。需针对算法决策设立相应的规制和救济机制, 注入公开、公平和责任理念, 进而兴利除弊, 促进科技向善和打造良性发展的智慧司法, 更好地实现“可视正义”。

算法只能延续和提炼人类知识, 但很难开拓和创造人类知识。算法决策能够通过程序性、公式化计算来优化诉讼程序和处理简单案件, 却难以处理重大的复杂疑难案件; 它能够促进形式正义, 却很难实现实质正义。算法有可能胜任法庭上的多项工作, 但无法胜任判决工作,<sup>[96]</sup> 包括证据标准指引、条件审查、校验纠错、裁判偏离度提示等, 都只能是人脑决策的辅助工具。关键之时还需要人来做决定, 由人来担任最终决策者。<sup>[97]</sup> 也许, 随着人工智能技术的突破升级, 算法决策替代人脑决策的空间会更大、质量也更高, 但它依然是司法助手、工作伙伴, 而不是独立裁判的“法官”。即便是“强人工智能”时代的“奇点”到来, 人类也会重新规划自身主体地位的发展蓝图。

---

**Abstract:** The judicial application of artificial intelligence (AI) technology is giving rise to the historical transition from “access to justice” to “visual justice”, leading to profound reform of the judicial operational mechanism, and reshaping the judicial process through recreated scene, digital coding of legal texts, modeling of judicial decision making and intelligent service management, thereby greatly improving the social efficiency of intelligent justice. However, the application of AI in the judicial realm has its inherent limits. For instance, algorithmic decision-making is not fully capable of absolute objectiveness and accuracy. When dealing with sophisticated and complex cases, it can promote procedural justice, but is faced with great difficulties in realizing substantive justice. In order to give full play to the function of AI in optimizing and reshaping the judicial process, it is necessary to regulate the use of AI and minimize its risks, establish regulatory and remedial mechanisms for algorithmic decision-making, introduce the ideas of openness, fairness and accountability, give full play to the advantages of technology and eliminate its harmful effect, promote the idea of “tech for social good” and the positive development of smart court, and better realize “visual justice”.

**Key Words:** judicial AI, visual justice, digital justice, smart justice

---

[96] 参见前引 [41], 多梅尔书, 第145页。

[97] 参见前引 [48], 巴赫曼等书, 第268页。