# 智能裁判系统的正当性追问

陈　巍

**摘要**：生成式人工智能技术的快速迭代显著提升了智能裁判文书生成系统的形式连贯性与逻辑合理性。人类法官认定事实需要自由心证，法律适用需要行使自由裁量权，个案裁判结果正确与否缺乏外在的客观实体性判断标准，现代法治通过“形式正义的法”原理解决法律的不确定性难题，裁判结果的正当性经由法律程序的公正性获得。AI法官生成的结果因生成式人工智能的特征而具有不确定性，同样缺乏外在的实体性判断标准，尽管学界借鉴传统诉讼的正当程序机理，提出技术性正当程序理论来规制算法决策，但AI法官的决策特征使得技术性正当程序机理无法发挥类似于传统诉讼正当程序的预期功能，AI法官的决策结果缺乏正当性基础，无法担起人类社会的价值决断者的重任，只能扮演人类法官的辅助者。

**关键词**：智能裁判；人工智能法官；形式正义；程序公正；正当性

**DOI**：10.19836/j.cnki.37-1100/c.2026.03.017

智能裁判系统[①]是人工智能技术应用于司法领域的一个重要场景，其依托法律法规与案例数据库，在输入的案件资料基础上自主提取分析案件关键信息、匹配法条，检索分析过往相似案例并自动生成裁判文书。当前，生成式人工智能以令人惊异的速度迭代，已有能力“自主”分析理解案件信息并快速生成一份形式上有理有据的裁判文书，可以做到以“假”乱真。随着技术持续迭代升级，这种能力会越来越强大。

当AI医生、AI画家、AI音乐家等表现出明显超越普通从业者的专业水准，以无可辩驳的优势击败一个又一个行业翘楚时，人类法官的裁判活动也会面临能否被AI替代的疑问。回答这种疑问需直面两个问题：人类法官所追求的司法公平正义到底是什么？AI法官是否有能力比人类法官更加出色地完成此种任务？如果回答是肯定的，那么随着技术的不断迭代进步，AI法官取代人类法官将是大势所趋，只是时间早晚的问题。因此，有必要从什么是司法公正这个原初问题入手，探讨智能裁判系统的功能和角色。

## 一、司法裁判的不确定性难题与应对机理

### （一）司法裁判的不确定性难题

在形式主义所处的“古典法律意识”时代，司法裁判被视为一个科学的、演绎的过程，但人们很快发现，期待法官经过推理就可以得出“唯一正确答案”不切实际。现实主义法学揭开了司法审判的神秘面纱，毫不隐讳地指出了司法自由裁量权的空间、原因及后果，揭示了司法审判不可消除的主观性，打破了人们对司法审判具有“唯一正确答案”的幻想。

法律从抽象文本走向个案裁判结果的过程不可避免地存在主观性，无论是事实认定还是法律适用，法官都有很大的自主选择空间。自由心证原则的确立，是绝对主义司法大坝上的第一道溃口，使得事实认定结论呈现出不确定的状态。法官对事实的认定依赖语言，而司法描绘案件事实的语言不可避免地具有模糊性。法律适用的核心是对案件事实作必要的法律性质判断，法官的眼光则需要往返于法规范以及案件事实之间[②]。为消弭法律解释的不确定性，具有实证主义倾向的法学方法学派

**作者简介**：陈巍，北京航空航天大学法学院副教授，北京科技创新中心研究基地副主任（北京 100191；ccww307@163.com）。

① 下文根据不同的写作场景和语境，有时使用“AI法官”指称“智能裁判系统”。本文中二者同义，不作进一步区分。

② 卡尔·拉伦茨：《法学方法论》，陈爱娥译，北京：商务印书馆，2003年，第13—14页。

发展出了一套系统的法律解释方法,试图用正确的解释方法达到正确的解释结果。但近乎繁琐的法律解释方法不仅不能弱化法官适用法律的主观性,反而在一定程度上加剧了法律适用结果的多样性。

在事实认定层面,法官需要判断证据与事实之间关联性的大小,此时法官依赖的是自己大脑中已有的经验法则,而生活经验无穷无尽,法律不可能预先规定。在法律适用层面,法官对某个事实是否符合特定法律规则的争议,看起来是由其对法律文本的理解分歧所致,本质上则是源于价值观的差异。不同解释背后代表不同的价值观偏好,也就是不同的利益优先级排序,法官对此依赖的是自己大脑中已有的偏好选择,必然存在个体性差异,并且法官裁判的主观思考过程仅在其大脑中发生,无法被外人知晓,具有"人脑黑箱"特征。正如波斯纳法官所言,"司法裁量权概念是一块空地或一个黑箱,当规则不够时,裁量权并不是解决如何判决案件问题的办法,而只是这个问题的名字。无论你把裁量权想象得多好,裁量权都会令法律职业界不安"①。

法官的自由裁量权长期以来神秘莫测,而司法系统始终强调司法裁判以法为据的确定性与客观性。近年来,人工智能技术通过对过往案例的大数据分析,揭示出法官鲜明的个人主观倾向。例如2016年,律师和机器学习专家迈克尔·贝内斯蒂分析了法国法官对庇护的决定,发现一些法官拒绝了几乎所有的庇护申请,而另一些法官通过了较多的申请。这项研究在法国引起了轰动,并导致一项立法将任何此类研究定为犯罪。法国新的司法改革法规定,有关法官或法院书记员的个人身份数据不得用于评估、分析或预测其实际或假定的专业实践的目的或结果②。这一禁止性规则可被理解为法律界为维持司法神秘感并捍卫司法权威的努力。

人类法官的裁判过程具有主观性,使得司法"同案不同判"现象一再出现。为实现法治所追求的稳定、可预期的社会秩序,这一不确定性难题显然必须被解决。从构建法治秩序的角度,所谓法律确定性是指将法律适用于案件时能够得出形式上唯一并且内容上正确的答案。如果能够得出这种"唯一的正确答案",那么法律的确定性难题也就能迎刃而解。如果把司法的目标设定为获得"正确的裁判",那么紧接着的问题是,个案中如何判断法官作出的裁判是正确的,即正确裁判的判断标准问题。司法实践中,大部分案件的事实认定与法律适用相对简单,争议不大,关于"什么是正确裁判"较容易达成共识。但一些案件在事实认定或者法律适用层面存在较大争议,不同观点各有理由,而不同裁判意见背后隐含了深层次的生活经验以及价值观层面的分歧。当个案出现不同的裁判意见时,并不存在一种"绝对正确的真理性结论"作为客观标准去评价判断各类裁判意见的对错,"真理符合论"在时间、空间和资源有限的司法场域并不适用。即便是最高人民法院作出的终局裁判,也不一定就是实体上正确无误的。因此,法律面临的不确定性难题难以从实体层面得到解决,法治秩序的构建不得不另寻他途。

(二)人类法官个案裁判的正当性机理

司法个案裁判结果的正确性缺乏客观的实体性判断标准,对此难题,现代法治通过"形式理性的法"加以解决。罗尔斯在《正义论》中提出了程序公正与实体公正的三种形态:纯粹的程序正义、完善的程序正义以及不完善的程序正义。诉讼程序通常被理解为一种不完善的程序正义,即程序之外的实体公正有其独立的判断标准。但是,由于人类认识能力有限,没有一种诉讼程序能总是毫无例外地实现实体公正③。司法实践中,当个案的实体公正陷入争议,不能任由争论无止尽地延续,因此不得不把经过公正程序的实体结果"拟制"为正义。由此,在理论上不完善的程序正义却在制度上作为纯粹的程序正义而发挥作用。审判结果是否正确无法完全借助外在的客观标准衡量,而充实和重视程

① 波斯纳:《法理学问题》,苏力译,北京:中国政法大学出版社,2002年,第27页。

② 施鹏鹏:《法国缘何禁止人工智能指引裁判》,https://www.spp.gov.cn/spp/llyj/201910/t20191030_436678.shtml,访问日期:2026年3月1日。

③ 约翰·罗尔斯:《正义论》,何怀宏、何包钢、廖申白译,北京:中国社会科学出版社,1988年,第80—83页。

序本身以保证结果能够被接受是其共同的精神实质。按照罗尔斯的分类来说，这里的倾向就是纯粹的程序正义。换言之，只要严格遵守正当程序，结果就被视为是合乎正义的。

形式理性的法律类型具有“纯形式的确定性”①。特殊的法的形式主义会使法的机构像一台技术上合理的机器那样运作。它把法律过程看作是和平解决利益纷争的一种特殊形式，它让利益纷争受固定的、信守不渝的“游戏规则”的约束②。在“形式理性的法秩序”中，关于程序的研究对理解法律至关重要，因为理性程序有助于以符合人类理性与自然法则的方式实施实证法律③。法律程序思想的根源在于现实主义法学对“实体权利和义务能通过法律推理实现自治”的挑战，法律程序理论基本上解决了现实主义法学提出的司法主观性的威胁，这种解决方案并不是贬损道德与政治哲学声誉的形而上学式的自说自话④。

在“形式理性的法秩序”中，人类法官对司法公正的追求，在一定程度上从追求个案实质正义转向追求经由公正程序的“合法性裁判”或“正当性裁判”。合法性被定义为一种普遍的感知或预设，即在一个社会构建的规范、价值观念、信仰和定义体系中，一个机构的行为被认为是可取的、适当的或恰如其分的⑤。当然，这种偏离并非放弃对实体公正的追求，而是将这种追求融入程序规则之中。正当程序是法院法治自信和司法公信力的主要渊源。手段的道德性逐渐包含合法性和正义的整体。实质正义是派生的，是无懈可击的方法的副产品⑥。随着程序理论不断发展，程序公正的内容越来越丰富，既有有助于实现实体公正的工具性价值，也有程序正义这种独立的程序价值，此外还融入了对于司法效率价值的追求。这些程序价值共同塑造了程序公正的面貌，也从不同方面巩固了裁判结果的正当性。

公正程序构建了一个约束人类法官的防护栏，法官在遵守公正程序的前提下，决定法律在现实世界的真正面貌，并在一些新型案件中发挥司法造法功能。形式正义的法在一定程度上解决了法律的不确定性难题，司法机关经由法定程序得以源源不断地作出一个个公正裁判，由此形成稳定有序的法治秩序。

## 二、智能裁判系统的机理与不确定性难题

### （一）智能裁判系统的机理

专家系统和机器学习是两种基本的人工智能技术路径。专家系统的知识通常以规则、条件语句和推理树等形式表示，使用基于规则的推理引擎，通过匹配规则和事实来推断结果。而机器学习依赖于数据驱动的方法，它不需要显式的规则或知识，而是从数据中学习规律和模式。包括深度学习在内的能够自行获取知识的各种机器学习算法，都是基于过去的数据并利用某种归纳偏好来预测未来的趋势。机器学习可以通过反复训练和新数据的输入进行更新，而无须人工手动干预，更适用于大规模数据集，可以自动从数据中学习并将结果应用于解决复杂问题。当收集和处理到足够多的过去事件的细节时，机器学习算法就会计算出这些不同的事件是如何相互关联的。

在事实认定层面，证据与待证事实之间的关联具有盖然性，证明标准制度本身就是概率论在司法领域的应用，对此算法具有独特优势。将统计学等数理逻辑通过AI法官引入司法证明，可以作为驯

① 马克思·韦伯：《论经济与社会中的法律》，张乃根译，北京：中国大百科全书出版社，1998年，第307页。

② 马克思·韦伯：《经济与社会》（下卷），林荣远译，北京：商务印书馆，1997年，第140页。

③ Eskridge W. N. Jr., “Metaprocedure”, *The Yale Law Journal*, 1989, 98(5), p. 964.

④ Fallon R. H. Jr., “Reflections on the Hart and Wechsler Paradigm”, *Vanderbilt Law Review*, 1994, 47(4), p. 970.

⑤ Suchman M. C., “Managing Legitimacy: Strategic and Institutional Approaches”, *The Academy of Management Review*, 1995, 20(3), pp. 574-575.

⑥ 诺内特、塞尔兹尼克：《转变中的法律与社会》，北京：中国政法大学出版社，1994年，第74页。

服不确定性的知识工具,其科学外观似乎可以消解事实认定主观性的“神秘主义”,让证明标准“可视化”①。将算法技术用于事实认定,可能干预甚至“篡夺”裁判者的事实认定权力②。在法律适用环节,一个案件的法律适用之所以产生争议焦点通常是因为,不同法官对某种事实是否符合某个法律规范有不同观点。算法分析可以通过统计过往案例中某种裁判意见的出现概率,找到概率最高的裁判意见。目前,我国各级法院正在尝试用人工智能技术挖掘利用海量司法案件资源,提供面向各类诉讼需求的相似案例推送、诉讼风险分析、诉讼结果预测。“同案不同判预警”“判决结果预测”“诉讼风险评估”等基于司法大数据的智能应用已经完成开发,进入应用阶段。

(二)智能裁判系统的显著优势

算法决策速度以秒计,可以24小时不间断地处理数量几乎不受限制的裁判文书生成任务,工作的处理规模甚至能超过全体人类法官的总和。除了效率优势外,算法裁判的准确性亦广受讨论。算法不受人类身体或精神状态的限制,如疲惫、压力或情绪等。相比之下,人类受制于所有可能的直觉扭曲,会过高估测现实,认为自己比实际中的更聪明,也存在情绪化的情形,人类的决定往往是带有偏见的、狭隘的、有倾向性的③,人类决策者的歧视也会损害人类主体的尊严④。即使是在崇尚公正的司法场域亦潜藏着不易察觉且难以消除的偏见,年龄、职业、性别等因素都会以微妙且隐蔽的方式影响法官决策⑤。而司法系统中的算法可能能够作出比人类更公平、更高效、更准确的决策,算法由于完全无视无关的主观因素,如一个人的穿着、他们在法庭上的行为方式等,因此不会像具有感性认识的人类一般被牵制。AI法官可能更便宜、更快捷,而且不易受到某些形式的偏见的影响,从而使法律制度不仅更加高效,而且更加公平,更容易被一部分诉讼当事人接受⑥。人类法官在裁判时,除了考虑案件的事实与法律问题,还会考虑有可能引发的社会舆论、对法官自身的潜在风险、上级部门的态度、败诉当事人作出过激暴力行为的概率等。这些考量往往与法律关系不大,法官不会公开表露,更不会在裁判文书中阐述,但可能会直接影响案件结果。AI法官则较难受到这些法律之外因素的影响。

人类社会培养一名优秀法官并不容易,短时间内不可能大规模培养。再优秀的人类法官一个人能处理的案件数量也非常有限,遑论还有许多专业技能平庸甚至不称职的人类法官。而算法之间的比较和竞争更为直接和高效,一种被广泛测试、验证和反复使用的先进算法,将快速淘汰其他劣质算法,可以被无限复制推广,适用于数以百万计的案件,这与优秀人类法官的稀缺性形成鲜明对比。一个高质量的智能裁判系统的研发成本或许极高,但数字技术特有的网络效应可以实现边际成本最低,这也是AI法官构想极具吸引力的优势。

(三)智能裁判结果实体正确标准的缺失

AI法官相比于人类法官具备显著优势,那么,人类法官面临的“实体正确判断标准缺失”的法律不确定性难题,对于AI法官生成的裁判意见是否不复存在?回答是否定的。

首先,不同智能裁判系统之间因算法设计和参数设定的差异,很可能在输入相同案件信息的情况下生成截然不同的裁判结果,“同案不同判”现象依然存在。这一现象的根源在于,人工智能模型通常基于开发者的偏好、训练数据的取舍以及算法优化目标的不同而存在差异,技术多样性使得不同AI系统生成的裁判结果无法保持一致,尤其当案件涉及较为复杂的法律解释或价值判断时,不同类型算法很可能“计算”出不同的裁判意见。

① 左卫民:《关于法律人工智能在中国运用前景的若干思考》,《清华法学》2018年第2期。

② 郑曦:《人工智能技术在司法裁判中的运用及规制》,《中外法学》2020年第3期。

③ 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、夏柯译,北京:中信出版社,2018年,第72、112页。

④ Levinson J. D., “Forgotten Racial Equality: Implicit Bias, Decisionmaking, and Misremembering”, *Duke Law Journal*, 2007, 57, p. 350.

⑤ 陈光中:《司法不公成因的科学探究》,《中国法律评论》2019年第4期。

⑥ Volokh E., “Chief Justice Robots”, *Duke Law Journal*, 2019, 68, p. 1140.

即便是全国法院使用同一智能裁判系统，其结果的稳定性也将受到生成式人工智能技术特点的限制。生成式人工智能的核心在于其深度学习模型的动态性，即根据输入数据和上下文不断调整输出。这种动态调整虽然增强了系统的学习能力和适应性，但这也导致AI法官在面对同一案件时，随着反复追问以及输入信息的变化而生成不同的裁判结论。这种“自我推翻”现象与司法追求的稳定性相违背。一旦智能裁判系统无法提供一个稳定的结论，其在法律实践中的权威性将难以确立。

其次，智能裁判系统依赖过往的海量案例进行大数据分析，对于新类型案件，AI法官可能因缺乏足够的既往案例支持而无法形成合理的裁判依据。人类法官面对新类型案件通常会结合法律原则、社会需求以及案件具体情节综合考量作出裁判，但AI法官只能依赖已知的数据进行推断。当面对前所未有的案件时，AI法官仅能基于概率倾向而非法律原则和社会需要输出某种观点，裁判结果的可信度和合法性均会受到质疑。

最后，即使AI法官能够通过大数据分析来发现过往裁判文书中的主流观点，这些“多数意见”是否真正代表当下的普遍民意仍然存疑。司法裁判不仅是当下社会主流价值观的反映，同时也是推动社会变革的重要工具。AI法官如果过于依赖历史数据中的多数意见，可能在价值观变化和社会需求演进的过程中表现出滞后性和保守性。同时，普遍民意本身并非一成不变，而是随着社会经济发展、文化观念更新等不断变化。今天的主流观点不一定代表未来的社会共识，因此对“历史意见”的过度依赖可能导致AI法官裁判结果无法适应社会发展需求，削弱司法对社会进步的引领作用。

从根本上说，法律的不确定性源于社会价值观的多元化与客观评价标准的缺失。特别是当社会处于巨大变革转型时期，不同价值观的碰撞冲突是正常现象。在搜索和概率统计方面，AI法官一定会比普通法官更强大，但AI法官生成的裁判文书仍然缺乏客观的实体正确评判标准，不足以从实体上让败诉一方当事人信服。

## 三、智能裁判的技术性正当程序理论及其局限

### （一）算法规制的技术性正当程序

学界关于算法治理的研究中，一种普遍的观点是通过程序性护栏，预防性地将算法决策纳入可控、可追责的轨道。“大数据正当程序”保障当事人在自动化决策程序面前依然能实质性地享有正当程序权利[①]。“技术性正当程序”针对算法的正当程序要求，建立了一个结构精心设计的纠问式质量控制模型，提供了能够增强嵌入自动化决策系统的规则的透明度、问责度和精确度的机制框架。常见的关于算法决策的程序要求，如透明度、算法影响评估、对决策结果的可解释权等，使不透明的算法系统更接近人类的决策机制[②]。这一思路显然借鉴了形式正义法治的正当程序机理，把针对人类法官的正当程序机制移植到算法规制领域，试图通过程序公正实现算法结果的正当性。

算法透明度原则大体可以对应审判公开原则，即要求算法的设计方或者使用方事前披露包括源代码、输入数据、输出结果在内的算法要素。这一原则可以通过算法备案、算法审查、算法第三方评估等程序性制度加以落实。算法的可解释权可对应程序参与原则、对审原则、辩论主义等程序制度[③]。算法解释权是技术性正当程序的核心，被视为“算法治理制度核心”[④]，试图在算法时代为人类掌控自身命运的能力保留必要余地。相对于算法透明原则，算法的可解释权具有更高的要求。

① Crawford K., Schultz J., “Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms”, *Boston College Law Review*, 2008, 55, pp. 124-128.

② Citron D. K., “Technological Due Process”, *Washington University Law Review*, 2008, 85, pp. 1305-1310.

③ 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。

④ Bambauer J., Zarsky T., “The Algorithm Game”, *Notre Dame Law Review*, 2018, 94, p. 36.

然而,落实算法的透明度原则容易出现一系列的困难,这是由于信息收集标准很难具有一致性,信息提供者和用户存在认知局限,以及涉及的内容可能具有敏感性。现实是,算法黑箱是由算法的技术特征造成的,而非人为刻意保持造成的。透明度作为一种观察、理解和管理复杂系统的方法,不仅是有限的,而且有时是具有误导性和无益的。认为算法决策可以用人类能够理解的语言"解释"清楚的观点也值得商榷。算法决策是令人费解的——它管理决策的规则相当复杂、繁多并且规则之间相互依赖,以至于无法从实践的角度对其加以考察。这些规则的意义也不在于从人类的角度被理解,而是要揭示和预测超出人类想象的精确模式及其相关性①。此外,算法作为企业的核心竞争力,通常被作为商业秘密进行保护,无论是基于劳动价值论,还是基于促进投资的功利主义原理,此举都具有一定的正当性基础。而应用于公共安全领域的算法与国家安全密切关联,也难以进行解释或实现透明化②。

为救济可能的错误,传统司法赋予当事人上诉的权利。算法治理也存在针对全自动算法决策的拒绝权制度。有关"不受完全自动决策约束的权利"被视为欧盟《通用数据保护条例》(General Data Protection Regulation,GDPR)中最神秘、最具吸引力、最富有远见的权利之一。但这种拒绝算法决策的权利可能仍然是一只"纸老虎"——理论上存在,但在实践中效用有限。即使是在GDPR生效几年后,关于拒绝权是什么样子的信息也很少③。此外,GDPR的拒绝权仅针对具有重大影响的"完全自动化"的决定而产生,也很容易因为表面上的"人工介入"而被规避。

(二)正当程序机理在智能裁判中的功能失效

传统法治针对人类法官的正当程序制度设计,至少具有如下四个方面的功能:首先是帮助人类法官更好地发现事实与准确适用法律的保障和促进功能,如法庭辩论制度确保法官有机会充分听取双方当事人陈述,防止偏颇片面,上诉制度通过让更有经验的法官进行复核,纠正一审可能的错误。其次是避免法官在行使自由裁量权时徇私、偏袒和恣意的防范功能,如回避制度、合议制度、审判公开制度。再次是通过程序正义彰显当事人主体地位并保障其人格尊严,由此获得当事人的信赖,巩固结果的正当性。最后还融入司法效率等其他价值的考量,如时限制度、失权制度等,实现司法公正与司法效率价值的平衡。但以上正当程序的多元功能面对AI法官几乎都是失效的,具言如下。

首先,技术性正当程序难以发挥智能裁判系统的"保障和促进正确结果"功能。对于普通当事人而言,算法裁判的机器学习技术机理过于复杂,很难真正实现算法透明和可解释,当事人也不太可能对智能裁判系统提出建设性的改进意见以提升其决策质量和准确性,只能被动接受算法决策结果。人类司法的二审和再审制度可以让经验更丰富的法官对一审结果进行复核纠错。但如果当事人对一审AI法官裁判不满,让同一AI"再算一次"或者让其他AI"重算一次",不仅不能缓和争议,反而会因为结果不同引起更大的质疑和混乱,毕竟算法之间的优劣不像人类法官的工作年限、资历、口碑一样容易辨别。其次,算法自身没有私欲可言,不会贪腐滥权谋取私利,那些可能影响人类法官决策的不可言明的地方性知识、潜规则等也没有存在空间。传统司法程序那些旨在避免人类武断决策并保证该决策可信与正当的经典制度安排,包括资质要求等声誉机制以及通知、参与、异议、救济等决策约束程序,在算法决策面前均告失灵④。人类司法程序的诸多制度设计对于算法决策而言是无效冗余的。再次,程序正义原则要求当事人富有意义地参与审判过程并有机会影响法官决策,而不能在被动的、无知的情况下接受后果。算法决策速度极快,当事人不可能介入算法决策过程施加影响,只能事后对结果表达异议,传统正当程序中关于尊重当事人主体地位的制度设计在算法决策中缺乏应用空间。最后,算法的决策速度

① Selbst A. D., Barocas S., "The Intuitive Appeal of Explainable Machines", *Fordham Law Review*, 2018, 87(3), pp. 1118-1122.

② 沈伟伟:《算法透明原则的迷思——算法规制理论的批判》,《环球法律评论》2019年第6期。

③ Kaminski M. E., "The Right to Explanation, Explained", *Berkeley Technology Law Journal*, 2019, 34(1), p. 197.

④ 解正山:《算法决策规制——以算法"解释权"为中心》,《现代法学》2020年第1期。

本身就是其显著优势，那些旨在促进效率的程序制度对算法而言缺乏实际意义。

更重要的是，司法公正的内容丰富，除了诉讼程序本身的公正之外，法官遴选制度、职业保障制度以及审判独立制度也发挥着极为重要的促进裁判结果正当化的功能。当事人之所以信任一个完全陌生的法官，允许其决定自己的重大利益，是因为他们相信，只有经过系统法学教育、严格考试遴选、长期业务考核等重重考验的法律专业人士，才有资格端坐在审判席上成为自己案件的审判法官。党的十八大以来，我国法官员额制改革进一步提升了法官的职业素质，这使得法官被视为最适合履行司法裁判职能的群体。审判独立制度也是现代司法的核心原则，确保法官能够不受外界干扰独立行使自由裁量权，否则通过正当程序约束法官的种种制度设计就会形同虚设。这一系列复杂的法律制度设计，使得败诉当事人面对人类法官作出的终审裁判，即便心有不甘也会选择息讼止争。与之相反，AI法官的裁判可能很难得到败诉一方当事人的真正认同，而坚持认为算法裁判是随机、偶然的产物，并非深思熟虑、专业严谨的裁判结果，因此不值得被尊重和信赖。

概言之，AI法官难以通过技术性正当程序机理解决生成结果的正当性问题。通过AI法官实现司法公平正义的构想将失去根基，成为无根的浮萍，虽可在水面上快速滋长蔓延呈现出繁茂景象，却始终无法扎根大地，以挺拔身姿担起缔造公平正义法治秩序之重任。

## 四、智能裁判系统的应用场景及限度

### （一）AI法官的应用场景

司法审判实践中，智能裁判系统至少有五种可能的应用场景：第一，人类法官已经有裁判结论，算法根据法官的提示和指示生成裁判文书，帮助其梳理和论证裁判思路，提高人类法官的裁判文书撰写效率。这种决策支持工具并不是用算法取代人类，而是朝着改进由人类决策者和机器辅助决策分析组成的社会技术系统迈出的又一步[①]。第二，法官在尚未有结论的情况下通过算法了解参考过往案例的多数意见，为自己的裁判提供参考或思路。如刑事审判中法官通过保释风险预测模型评估风险，以此作为判断罪犯能否被假释的重要参考[②]。当然，法官也可以在斟酌考虑后不采纳算法的意见。第三，算法生成的裁判意见，人类法官有权采纳，也有权拒绝，但需要说明不予采纳的理由，否则可能引起一定的不利后果。常见方式是通过司法系统统一部署的算法形成裁判预测，再将法官的裁判结果与预测结果进行比对，发现两者的“偏离度”，法官需要回应并解释其判决的“偏离度”，以实现同案同判的目标。这种应用场景下，AI法官将对人类法官产生一定的约束效果。第四，对案件进行分类，在简单案件中，AI法官的裁判可以获得一定法律效力，提升此类案件的审判效率，而疑难复杂案件则仍然由人类法官裁判。第五，算法自主生成的裁判结论具有强制性的法律效力，AI法官事实上取代人类法官决策。有观点认为，人工智能会改变并代替法官裁判，机器裁判超越法官裁判的“奇点”必然会到来[③]。

第五种应用场景前已论述，AI法官既无法从实体公正层面获得裁判正确性的判断标准，也缺乏经由正当程序获得的裁判正当性基础。即便算法如何先进，甚至具备优于普通人类法官的价值衡量与说理能力，也因为缺乏正当性基础而不能成为独立的法律秩序裁判者。当前学界所热议的针对司法人工智能应用的技术性正当程序理论，其初衷是借鉴传统正当程序机理提升智能系统裁判的正当性，但此种思路忽略了人类法官与AI法官的实质性差异，前景值得怀疑。

第四种应用场景的问题在于，如果是简单案件，法官很容易作出有共识的正确裁判，不需要借助

① Balkin J. M.,“The Path of Robotics Law”, *California Law Review Circuit*, 2015, 6, pp. 48-49.

② Moses L. B., Chan J., “Using Big Data for Legal and Law Enforcement Decisions: Testing the New Tools”, *University of New South Wales Law Journal*, 2014, 37(2), pp. 660-663.

③ 马皑、宋业臻：《人工智能“法官”的一种实现路径及其理论思考》，《江苏行政学院学报》2019年第2期。

AI形成裁判意见。简单案件实现司法效率的方式很多,也可以选择利用人工智能技术的优势提高法官裁判的效率,而没有必要直接赋予AI法官的裁判法律效力。况且简单案件与复杂案件的区分在实践中也较为模糊,容易引起规则适用的分歧和混乱。

第三种应用场景是一种很有吸引力的方案,既能尊重人类法官的独立性和自主性,也能实现AI法官在实现同案同判或类案同判方面的约束性功能。但此种方案也有较大的风险。虽然要求法官对其意见与算法预测结果之间的"偏离度"作出解释的初衷是提高裁判的透明度和一致性,但实际上却可能加重法官的工作负担并增加其心理压力。在当前高强度的司法工作环境下,如果公众和上级法院将算法生成的意见视为"标准答案",法官的偏离很容易被视为"异常"甚至"错误",这很可能导致法官在裁判过程中采取防御性态度,为避免解释负担和争议而趋于保守,简单接受智能裁判系统生成的文本。这种防御性态度不仅可能抑制法官对复杂案件的深入思考,也可能限制司法裁判的创新引领功能,在新类型案件和复杂价值冲突案件中更是如此。更严重的是,这种机制可能让法官更多地关注如何满足算法预测,而非着眼于个案的事实和法律,最终损害司法的独立性和灵活性,使裁判流于表面形式,难以实现真正的司法公正。同案不同判是法治无法彻底消除的现象,我们需要尽可能降低其负面影响,在常规性案件中实现裁判的一致性与可预期性,但这种现象蕴含的司法制度创新功能与保持法律与时俱进的益处,也是一种值得珍惜的司法价值。

第一种和第二种应用场景值得肯定与推广。当前,学界主流观点是AI无法取代人类法官,但这一观点的论证往往是基于AI技术本身的缺陷,即其生成裁判文书的质量堪忧,不足以让当事人以及公众信服。可以预见,人工智能技术的功能将愈发强大,智能系统生成的裁判文书也将越来越逼近优秀的人类法官裁判文书的水平,特别是其说理部分,表现出来的专业水准甚至会高于一部分人类法官。但无论技术多么先进,AI都只能扮演人类法官工作助手的角色。人类法官可以自由使用AI,哪怕原文照搬了AI生成的文本,只要其在对外公布的裁判文书上落款署名,就将被视为人类法官的裁判意见产生相应法律效力,并由人类法官对文书内容负责。

(二)司法责任制框架下的司法AI应用

生成式人工智能的"幻觉"问题是当前智能司法最受关注的技术缺陷之一。AI生成的内容可能无中生有地援引并不存在的案例、捏造法律条文或曲解法律规范,且这类错误往往包裹在看似严谨流畅的表述之中,不太容易被识别。一旦此类错误未经甄别地进入裁判文书,不仅会损害当事人的权益,也会动摇司法公信力的根基。因此,无论法官在多大程度上使用智能裁判系统,其对AI生成的内容进行认真负责的复核和甄别,始终是不可替代的责任。事实上,优秀的法官能够以深厚的法律素养和职业经验识别并纠正算法的偏差与错误。

从制度设计的角度看,规范法官使用AI的最有效路径,并非对具体使用方式加以细致规定,而是通过司法责任制这一既有制度框架加以统摄。司法责任制的核心逻辑在于,法官对其署名的裁判文书承担完整的法律责任,无论该文书的生成过程中借助了何种技术工具,最终的法律后果均由署名法官独立承担。法官不能以"系统错误"为由推卸责任,对AI生成内容的审核义务是其履职尽责的应有之义,法律无须制定特别规则让算法提供者分担法官责任。

在此框架下,法律对于法官个人如何选择及运用AI工具,不宜进行过度干预。法官的专业判断能力、工作风格与知识结构存在个体差异,对AI的依赖程度与使用方式因人而异本属正常。过度的法律规制不仅难以实现实质性监管效果,还可能会无益地限制法官在技术应用层面的探索与创新。司法责任制的内在激励机制自会促使法官在使用AI时保持必要的审慎与独立判断,而这正是任何外部规制难以替代的自律动力。

人类法官的裁判远非完美,但司法裁判的本质不同于自然科学对真理的探求,其根植于人性、道德与人类社会基本价值选择,是一种人为的技艺。相对于瞬间生成结果的算法,那些显得保守、冗长、繁琐的司法制度,恰恰蕴含了法律秩序得以获得公众认同以及令败诉者服判息诉的深层理据。越来

越“聪明”的AI是一面镜子，虽映照出了人类法官裁判制度的不完美，但也凸显出人类构建的司法制度的厚重与精妙。人们可以对一位优秀人类法官产生发自内心的尊重与敬意，但很难会对一种算法产生尊崇感；人类法官可以在争议案件中作出前瞻性判决，打破旧时代桎梏、引领时代新方向，而一份算法生成的裁判文书，哪怕内容上并无二致，也无法获得类似的权威。民众对司法的信任，与其说是对人类法官作出的裁判文书正确性的信任，不如说是对法官能力和品行的信任，是对由法官所代表的国家整体司法制度的信任，而人工智能技术无法企及。

## 五、结语

当前，人工智能技术在传统司法领域的应用仍只是对传统法律实施流程的优化，并非颠覆式创新，例如法条与案例自动检索、法律文件自动审校和自动生成、案件结果预测等。事实上，技术对于司法制度更深层次的冲击是通过技术实现法律规则的自动实施，从源头上避免纠纷发生，让法院“无案可判”。例如，未来自动驾驶技术的全面普及或将使得因人类驾驶机动车引发的交通事故损害赔偿、酒驾、醉驾等民事、行政和刑事案件基本消失。这种科学化、非道德化的机器规制，或将完全并直接取代法律的规范功能，规范性期望会被认知性期望取代[①]。在前数字时代，事实真相发现困难，法官探索出了许多富有智慧的事实认定经验和技巧，被视为一种宝贵的司法经验。进入高层级的数字化、智能化时代，事实认定可能就将由系统直接调取数据进行自动识别处理，传统司法经验可能成为屠龙之术而无用武之地。

智能法治是人类法治文明史的一次质的飞跃。当然，这一构想成为现实并非易事，还需要很长一段时间的持续探索。法律规则即使被编码成可自动实施的算法，仍需要法律人来决策法律编码算法的具体内容并不断完善，算法无法处理的疑难复杂问题也需法律人予以解决。只要人类社会存在对法律公平正义的需求，人类法官就依然还有大展身手的广阔空间。

## Questioning the Legitimacy of the Intelligent Adjudication System

Chen Wei[1,2]

（1. School of Law, Beihang University, Beijing 100191, P.R.China;
2. Beijing Science and Technology Innovation Research Center, Beijing 101117, P.R.China）

**Abstract**: The rapid development of generative AI has brought intelligent adjudication systems increasingly close to human judges in terms of formal coherence and logical soundness. Whether AI judges can genuinely replace human judges depends on two prior questions: what gives human judicial decisions their legitimacy, and whether AI-generated rulings can establish a comparably convincing foundation for such legitimacy.

Judicial adjudication is inherently subjective, both in fact-finding and in the application of law. Yet there is no external, objective standard against which the correctness of any individual verdict can be measured. Confronted with the problem of legal uncertainty, modern western rule-of-law systems turned not to an endless pursuit of substantive truth but to the principle of “formally rational law.” The underlying idea is that a decision reached through fair procedure is to be regarded as

① 余成峰：《法律的“死亡”：人工智能时代的法律功能危机》，《华东政法大学学报》2018年第2期。

legitimate. Procedural fairness thereby became the primary source of judicial legitimacy. The due process serves multiple functions: promoting accurate outcomes, constraining judicial arbitrariness, safeguarding the dignity of the parties, and advancing the efficiency of justice. Together, these functions underpin the acceptability of the human judicial system.

AI judges face a very similar challenge of legal uncertainty. Different intelligent adjudication systems, varying in algorithmic design and training data, may reach diametrically opposite conclusions on identical facts. The dynamic nature of generative AI can also produce contradictory outputs from the same system when prompted with only minor variations in input. When confronted with novel cases lacking adequate historical precedent, an AI judge can do no more than follow statistical tendencies rather than reason from legal principle and social need. These features confirm that AI-generated judgments are equally without an external, objective standard of substantive correctness.

Scholars have attempted to resolve this problem by adapting the logic of human due process, proposing a theory of "technological due process" that would subject algorithmic decision-making to requirements of transparency, explainability, and the right to contest automated outcomes. However, the difficulties are fundamental. The core functions of human due process—facilitating fact-finding, checking judicial arbitrariness, ensuring meaningful participation by the parties, and promoting efficiency—are largely inapplicable to AI adjudication. Algorithmic opacity makes the transparency principle exceedingly difficult to implement. Because algorithms have no personal interests, and procedural safeguards designed to address individual bias, such as recusal and collegiate deliberation, lose their practical significance. The speed of algorithmic decision-making leaves parties no meaningful opportunity to intervene in and shape the outcome. Appellate review mechanisms can't reach the ultimate answer. Beyond procedural considerations, the legitimacy of human judges also rests on rigorous professional selection, career tenure protections, and the principle of judicial independence—an institutional architecture that leads losing parties, however reluctantly, to accept final judgments. This is an authority that AI judges plainly cannot command.

It follows that, however far technology advances, AI judges can function only as assistants operating under human direction. Judges may freely draw on AI in preparing their decisions, but the moment a judge signs a verdict, its legal force and legitimacy vest in the human judge. Judicial adjudication is grounded in judgments about human nature, moral values, and the needs of society. No algorithm, however sophisticated, can replace it.

**Keywords**: Intelligent adjudication; AI judges; Formal justice; Procedural fairness; Legitimacy

[责任编辑:苏　捷]