
人工智能于法官绩效考核之应用

——以程序法视角为中心

梁庆¹ 韩立收² 刘信言³ 郑儒传⁴¹

(1. 南京大学 法学院, 江苏 南京 210093;

2. 海南大学 法学院, 海南 海口 570228;

3. 中国人民大学 法学院, 北京 100872;

4. 海口市人民检察院第二检察部, 海南 海口 570125)

【摘要】: 人工智能可用于评估司法裁判, 取代以结案率为主的法官绩效考核方式, 这一过程实质是运用人工智能完成法律论证。法律论证无疑需要经验, 但是法律推理本身并不需要经验, 并且人工智能通过观察学习案例和不断演练裁判可以建立与人类相似的司法经验。运用人工智能绩效考核的行政属性决定其并未突破法官独立审判的制度框架。与此同时, 人工智能通过建构算法模型, 预先已经将归属法官自由裁量的部分排除在外, 与自由裁量权亦不冲突。总之, 人工智能不仅能够克服既有法官绩效考核的缺陷、给予法官自由心证必要约束和辅助、间接规范我国诉讼使之同案同判, 还有助于提升司法裁判公信力。

【关键词】: 人工智能 裁判评估 法官绩效考核 规范诉讼 大数据

【中图分类号】 DF7, DF8 **【文献标识码】** A **【文章编号】** 1673-0186 (2021) 011-0078-013

近年来, 虽然学界与实务界对于人工智能的法律应用这一法学前沿问题已经讨论甚多¹, 且关注侧重点各有不同(学界主要关注人工智能的主体资格与法律责任以及人工智能创作物的著作权归属问题, 实务界则更关心如何利用人工智能辅助律师高效工作、能否快速准确地预测案件结果), 不过, 其仍有一项重大用途被完全忽略, 即人工智能可以用来评估我国诉讼裁判之全过程。

一、问题之提出

一国司法水平的高低, 由成千上万个案件的诉讼质量之合力决定。诉讼行为本质是人的行为, 最终决定每场官司成败的是法

¹**作者简介:** 梁庆, 南京大学法学院博士研究生, 研究方向: 民事诉讼法; 韩立收, 海南大学法学院教授, 研究方向: 法学理论; 刘信言, 中国人民大学法学院博士研究生, 研究方向: 法学理论; 郑儒传, 海口市人民检察院第二检察部检察官助理, 研究方向: 刑法学。

基金项目: 国家社会科学基金项目 (19ZDA347)

官裁判决策之优劣。而使法官群体努力作出公平正义的审判离不开奖惩分明，奖惩分明又取决于法官裁判的绩效考核评估方法之得当。不过，当下法官绩效考核方法固有的缺陷（偏重结案率）导致其经常“好心办坏事”。而不当的激励措施又反过来使法官作出倾向于更能提高个人绩效的审判决策（而不是更公正的判决），形成恶性循环，使公平正义之原初目的最终适得其反。

庆幸的是，在渐行渐近的人工智能时代，或将有有望改良现行的法官考核方法。如同前三次工业革命一样，新技术并未对法学理论本身形成太多实质挑战^{[1][42]}，而是悄悄改变司法的过程（如测谎仪的使用）²。只不过，人工智能技术所产生的变革更加猛烈。

一方面，各国研发人工智能的初心是要全面模拟人的智能以替代人类劳动，如无人驾驶汽车；另一方面，在一些特殊领域，如人工智能法律系统能否代替法官，不仅是技术问题，更是政治问题^{[2][37]}。然而，本文并不打算对此争议展开讨论，而是旨在探究运用同一技术（人工智能司法论证系统）对司法裁判进行评估的可行性，用其取代目前偏重法官结案率的绩效考核制度，通过人工智能技术客观地评估裁判，间接规范法官行使裁判权，进而整体规范我国诉讼裁判活动。

应用人工智能考核法官裁判至少有三个层次的问题待解决：首先，“人工智能”可能是计算机技术中被当今社会滥用程度最高的术语之一，有必要对于人工智能的概念在法律层面上的内涵予以清晰的界定，在此之后才能进一步讨论用其对法官展开绩效考核是否具有现实基础；其次，相较于人类，应用人工智能进行绩效考核优势何在，需要阐明；再者，还必须知晓使用人工智能技术评估司法裁判的缺陷所在，并加以解决。

二、人工智能应用于法官考核的现实基础

虽然“人工智能”的定义本身已经形成共识，但是其在法律层面的定义直到现在仍不清晰。学术讨论必须借助精准的概念作为先决条件，否则争论下去只会无的放矢、徒费口舌。因此，完全有必要将人工智能之概念于法律层面所表示的内涵予以阐释、界定在先。

（一）法律层面的人工智能定义

人工智能被定义为“计算机科学的一个领域，它涉及给予机器看起来像人类智能的能力，并具有模仿人类智能行为的能力”^[3]。20世纪80年代初，美国学者约翰·R. 塞尔（John R. Searle）所提出的“强人工智能”与“弱人工智能”的概念在世界范围内具有较大影响^[4]。

在此之后，新西兰学者又进一步提出了强人工智能、认知模拟（相当于弱人工智能）、应用人工智能的分类^[5]，这也是眼下人工智能研究的主流看法。另外，近年来牛津大学人类未来研究院的学者设想的超级智能也影响甚广^[6]。

对于人工智能在法律层面的定义，布鲁斯·G. 布坎南（Bruce G. Buchanan）和托马斯·E. 海德里克（Thomas E. Headrick）在1970年发表的一篇文章是人工智能在法律领域最早的讨论之一，他们写道，人工智能必须反映一个人的推理过程。具体来说，人工智能必须模仿律师的分析推理才能达成法律解决方案，法学院课程中的法律研究和写作技巧需要纳入计算机程序^[7]。

1990年，埃德温娜·里斯兰（Edwina L. Rissland）在《耶鲁法律评论》发表了另外一篇关于人工智能的开创性论文，她指出人工智能在法律层面的目标是：首先能够理解规则、案例和论据等基础知识，然后利用论据和先例进行推理，最终完成法律解释与论证^[8]。人工智能的一系列概念在法律层面上的对应情形，可以归纳如下：

第一，应用人工智能：如指纹识别、无人驾驶等，主要是指拥有简单的逻辑和计算能力，能够商用化的智能系统。这在法律领域已经基本实现，包括赵廷光教授的《实用刑法专家系统》，武侠律师的“法律机器人”，伦敦大学开发的“人工智能法官”系统^[2]，以及全国各地陆续推出的“智慧法院系统”³。

第二，弱人工智能：如阿尔法狗，指可以模仿人类的神经网络进行复杂推理和深度学习，具有非逻辑思维的能力，但思维过程仍然受到人类为其设置的框架结构制约，且无法感知周围环境。在法律层面意味着，人工智能可以像法官或律师一样完成法律推理及解释，也就是说，其完成法律论证仍然是按照人类法律专家的思维进行的。迄今为止，掣肘这一技术成为现实的关键一步是机器对于人类语言的理解。

第三，强人工智能：是指各方面都能和人类比肩，思维过程不受人类为其设置的框架制约，能够感知周围环境，可以产生和人类不一样的观点、结论。可见，如果使用强人工智能作为法官，人工智能很可能运用不同于人类的“自己的方法”来解释、适用法律。这样必然也会造成一定程度的不可控。

第四，超级智能：不仅能感知周围环境，独立于人类思考，且在几乎所有领域远远超过人类，从而会给人类世界的存在造成威胁。这一完全失控的局面在法律层面就意味着，机器将以机器自身创造的法律理论来审判人类。

（二）实现司法论证的技术过程

以人工智能代替法官裁判，与用人工智能评估法官裁判，事实上是同一技术过程，即应用弱人工智能进行司法论证的技术。本文认为，弱人工智能可以完成司法论证，且这一技术极具现实性基础。

1. 人工智能进行司法论证的步骤

既然全部法学研究皆是围绕法律论证（立法论证和司法论证）进行，那么法律实践中人工智能应用之目的当然也是围绕司法论证而展开。对于简单案件，涉及规范单一、事实清楚，上文中提到的应用人工智能可以完成论证。这些人工智能法律系统的特点是，没有对证成阶段的证据推理和法律适用阶段的法律解释进行模拟，而是直接把案件事实当作法律推理小前提，用三段论演绎推理方法得出裁判结论。

而对于复杂疑难案件，由于作为三段论内部推理前提的小前提法律事实和大前提规则，本身含糊不清，需要先进行外部论证。瑞典法学家瓦尔格伦（Peter Wahlgren）从人工智能角度对复杂疑难案件的法律推理过程进行微观研究，将其描述为7个阶段，包括：证据推理、法律检索、解释、规则适用、评价、学习、简述^{[2]27}。

当下，这一过程瓶颈在于复杂的证据推理和法律解释。另外，美国在部分法律领域使用的人工智能系统，包括合同分析法律科技公司和专利纠纷案件判决结果预测系统（Lex Machina）^[7]，并非是人工智能司法论证技术研发成功，而是基于大数据的简单分析预测。

2. 实现证据推理与法律解释指日可待

有学者认为，人工智能无法实现对证据的取舍和证明力大小的判断。理由是“人工智能是线性的，不具有思维的跳跃性和创造性。因此，人工智能只能对海量证据进行初步筛查、提取和判断，却无法做到证据的深加工和有机组合；更不用说达到法官依据审判经验对证据证明力的有无和大小的判断和取舍”^{[8]113}。

同意这一观点的学者，进一步提出人工智能用于证据判断应遵循辅助性原则^{[9]62-63}。不过，这两位学者混淆了证据推理与司法论证所用方法的边界。司法论证无疑需要经验，法律推理（包括证据推理）本身并不需要经验。既然是推理，就可以通过“If...then...”的规则处理^[10]。

这种“思维方式”的过程可以通过严格的规则体系来实现^[11]。至于应用人工智能进行法律解释能否研发成功，质疑主要集

中在人工智能没有七情六欲和价值观，而法官解释法律需要考虑人情和良知^{[2]34}。不过，现代法治观念中，对于人治和法治的区分，本就将法官设想成机器（概念法学的经典命题：法官自动售货机论）^[12]。

在美国，只有最高法院的案件审理才会以政策（价值观）为主^[7]。所以，这一问题实则为伪命题。实际上，运用人工智能证据推理和法律解释，真正困难的是如何把法律推理中已经存在的若干种知识（前提）精确地表述为机器语言。“语言问题”是人和机器相互适应的一个过程，需要实践的缓慢积累。不过，可喜的是，在如何教机器学会理解人类语言文本这方面，近年来在科研上取得了突破性的进展⁴。

3. 不仅包括逻辑，同样蕴含“经验”

“法律的生命不在于逻辑，而在于经验”，霍姆斯大法官这句名言可以说是很多人拿来质问人工智能推崇者的最有力武器。不过，他们这样认为的前提是人工智能只有逻辑、没有“经验”^{[8]114}。但正是这一前提存在问题。弱人工智能与应用人工智能的区别就在于，它不是一个封闭的系统，而是通过深度学习，摄取“经验”来不断完善自身。

以名噪一时的柯洁与阿尔法狗人机大战来看，人工智能通过学习数以万计人类专业选手的现存棋谱以及自我对弈提升棋艺^[13]。国外学者也指出，人工智能并非一定依赖于使用形式逻辑作为推理工具，其可以通过案例训练从而产生“经验”^[11]。易言之，在法律层面，人工智能通过观察学习案例和不断演练裁判不仅可以建立与人类相似的“司法经验”，而且积累的速度和数量级是人类不可与之匹敌的。

（三）道德与政治的双重合法性

人工智能司法建立在庞大的司法数据之上，可以为绝大多数案件提供作为参考的先例、标准和刻度。只要承认特定社会生活的历史连续性和生产性，而不过分强调社会进化的突变性，人工智能司法就符合社会演进规律和法律信念^{[14]34-35}。这种观点从大数据是道德与伦理观念之反映出发，指出人工智能司法的道德合法性。

不仅如此，由于大数据所能搜集的法律信息更为广泛，相较于人类来说，对于某一案件的评估更为公允。例如部门法之间有可能发生冲突，不同的指导案例之间也完全可能发生矛盾。利用人工智能对于大规模数据的分析对比，则很容易发现其中的不同之处，从而作出客观中立的评估。而这种对于法官裁判不偏不倚的评估考核，本身又构成了一种社会公意的执行。

有观点指出，司法人工智能兴盛的起因是市场与官方的双重驱动。一方面，我国司法资源严重不足，司法人工智能的需求被激发出来，非官方的公司与律所出于市场利益投入研发。另一方面，最高层意识到人工智能的战略意义之后，在政法体制下，中央的布置与顶级司法系统的接受，以及地方的迎合，形成共识。

国家政治系统的发动迅速转变为司法系统的核心任务^[15]。笔者同意这种观点，并且认为，人工智能究其本质是一种世界性的、革命性的技术，法学界既不能视若无睹，也不必谈虎色变，而是应该提前布局、抢占先机，在当前我国自上而下推行试验人工智能之际，乘天时地利人和之便，紧抓机遇、进步司法。

三、人工智能应用于法官考核之优势所在

法官绩效考核本质是，通过复核案件相关事实的认定以及法律适用情况来评价审判质量。人工考核劣势在于，考核者本身对于每起案件的事实认定与法律适用难以维系标准的统一，存在过多主观性，而评估案件的客观性恰恰也是人工智能的优势所在。

（一）克服既有法官考核与审级制度之缺陷

1. 摒除既有法官考核制度之弊病

“当法官们为着完成那些诸如结案率、上诉率、发改率、撤诉率等各种指标而作出裁判的时候，他们就不可能只服从法律和良心，他便有可能为这些指标而迁就某些力量和因素。”^[16]¹³⁻¹⁴

以结案率考核法官之弊病是一个饱受诟病的老问题，并不新鲜。现实当中，各地法院还常常结合上访率辅助考核（纯粹认为只要结案率高、上访率低等于法官裁判质量高）。不过，笔者仍然想从因果性和相关性两方面提出逻辑上的批判。

其一，当事人是否采取上访和案件裁判结果公正没有因果关系，即使案件审判错误、存在不公，当事人未必上访；案件裁判正确公正，当事人也可能上访。

其二，退一步说，哪怕上访率与错案率有微不足道的间接关系，这种考核方式仍然受到截访这一普遍现象的干扰，上访率低可能完全是拦访截访的结果，相关性大打折扣。近些年以上海、重庆、武汉、南京为首的国内一些中级人民法院，也开始探索实践考核方式多元化。

其中，重庆一中院和武汉中院走的是两条不同的路，前者研发了复杂维度案件质效分析系统，在考核方法上以定性考核为主、定量考核为辅；后者设计了庞杂的量化考核体系，由一级指标、二级指标总共数十个指标组成^[17]¹²⁴。

不过，单纯量化考核的方式应用在评价司法行为上是弊大于利的。这种考核方式至少用在考核司法行为和学术造诣上，是弊大于利的。但是在方法上，利用人工智能模拟司法论证来考核法官相对来说客观公平。这是因为，以司法论证来评估司法论证，无疑是最佳方式。

2. 解决既有法院审级制度之悖论

审级制度设计的初衷，无疑是在一审错判时能实现及时发现并纠错之功能。对于这一点，国内外的教科书都鲜有质疑。然而，只要我们承认两个常识，便会发现这一制度的效果几何，并非完全不可怀疑。

一是，存在第一审裁判正确，第二审改判错误的情况。审级制度的悖论在于：多年来我国二审案件 90%以上维持原判^[18]，对于不到 10%的改判案件，“错改对”是否多于“对改错”，在实证上很难研究。于是，审级制度纠错之功能，多半是人们出于希望公平正义多一份保障的臆想。二审真正的意义在于阻隔当事人不满判决或败诉以后极端情绪的瞬间爆发，使之通过二次审理的心理适应期慢慢接受不满诉讼之现实。或有人质疑，我国二审采取续审主义观点，允许二审提交新证据材料。但是，更多信息，同时意味着更多虚假信息，法官就面临更多的判断。

二是，虽然中院案件少，但是二审审限更短，二审法官并不因此拥有比一审更充分的时间来审判。

三是，通常认为审级制度能起到监督的作用，易言之，由于当事人申请二审有可能推翻一审，故原审才会谨慎为之。然而，恰恰由于二审 90%以上维持原判这一大数据的公开，反过来给一审法官吃了定心丸，令其相信一审被推翻只是小概率事件。

3. 以裁判之道评估裁判

用人工智能评估法官裁判的本质是：以人工智能作出的司法论证评估法官的司法论证，或者说以人工智能裁判评估人类裁判，是一种直接评估方法。如此，它便比以结案率为主的间接评估方法可靠。这种可靠源自司法行为与行政行为性质上的区别。

对于行政绩效的考核，通过量化其履行程度来评估是可靠的，因行政行为的性质本就是按时按量执行政治任务。但司法行为的性质是判断，若是用量化的方法考核裁判绩效，相当于不论判断对错，只要做出判断的数量多，绩效便越高，这显然荒诞无稽。

审级制度实际上是两个级别法官之间的判断效力问题，与对错无关，唯一确定的是高级别法官之判断最终生效。但上文已述，生效的未必是正确的。多数情况下，这两个级别的法官，业务能力、时间充裕度皆相差无几，其做出的又都是主观判断，恐难说究竟谁对谁错。而人工智能做出的司法论证具有客观性，二审法官以人工智能对一审法官判断之分析判断作为参照，具备主客观结合之优势。

（二）规范诉讼，同案同判

1. 对于诉讼程序的监督规范

依托于大数据和快速分析的优势，人工智能在评估诉讼程序的得失方面大有用武之地。以立案阶段为例，目前有许多复杂的情况造成重复起诉问题。比如，原告公司P起诉其员工D借钱不还，唯一的证据是一张收条。因收条不能证明债务关系，法院判决P败诉。之后P又以不当得利起诉，同一法院判决P胜诉。

同一法院，同一事实，前后裁判结果竟然完全相反，这显然是不合理的！不过，倒也不能全怪立案庭的法官，这里面有立法上关于“诉讼标的”界定模糊的问题，有举证责任的问题，也有既判力的问题，可以说十分庞杂。况且，处理前后案的可能不是同一法官。再者，即便是同一法官受理，人脑毕竟不是电脑，在案多人少的大环境下，面对各种复杂的案情，实属强人所难。

可人工智能却很容易瞬间识别出哪些案件涉及重复起诉。这还只是立案阶段，在评估一审、二审的常规性程序、诉讼标的识别、证据规则、既判力的时间界限以及程序法与实体法的交叉领域，人工智能评估系统同样大有可为。从而可实现整体上评估我国案件程序对错与否，保证法律统一适用、同案同判。

2. 对于法律适用的监督规范

人工智能评估系统除了可以发现案件程序错误以外，还能够有效评估法律适用，法律适用错误在刑事诉讼和民事诉讼中都极为常见。在刑法实务中，有为数不少的正当防卫案件原本符合无罪成立要件，实务中却被认定为犯罪。检索案例库，类似“昆山反杀案”“涞源反杀案”等案件，在以往被定性为正当防卫的寥寥无几。

主要原因有两方面：一方面正当防卫的适用条件拿捏不好；另一方面是怕施害一方上访，为平衡双方利益，不惜违背正当防卫立法之精神，将正当防卫一方入罪，同时从轻处理。结果常常是行政干预的做法占据主导地位。通过铁面无私、没有“人情”的人工智能来对案件进行评估，就不难发现“和稀泥”的问题。

司法实践中，有时极为简单的法律关系判断也可能出错。例如有法官将单位丢失员工档案的纠纷（侵权法律关系）误判为劳动争议关系，认为丢失档案行为不由侵权实体法规范。总之，用人工智能评估法官裁判，可以协助法院发现案件中的法律适用错误。

3. 对于参照指导案例的监督规范

首先，对待指导案例的态度，本文认为应该尽力遵循，哪怕指导案例是错的。如果指导案例本身存在错误，因其对所有人公开，故法学家、律师、甚至法官群体自身早晚能够发现错误之处，然后各自可凭借公开质疑（如学术期刊的发表、社会新闻舆论），促使指导案例编选者通过“顶层设计”取舍抉择。

可如果置指导案例于不顾，势必造成法律适用不统一、同案不同判之后果。实务中，法官常以手中案件与指导案例不完全一样而拒绝参照。可是这个理由却站不住脚，全世界没有两个完全一模一样的案件。如果以两个案件完全一样为标准决定参照与否，那世界上就不会有判例法国家的存在，我国指导案例制度也就全然失去了意义。

时至今日，人工智能类案推送功能早已在现实中使用，就是前文所说的应用人工智能。不过，这还不够，弱人工智能可以进一步评估法官对于指导案例的参照与遵循情况。笔者以为，人工智能运用指导案例评估法官裁判的工作原理可以设计为：首先，通过对于大量案件的筛选，找出案件事实与该案事实基本相符（可供参照）的指导案例；然后，通过自身司法论证功能，推演该案结果是否与指导案例结果一致，如果一致，则证明该指导案例很大程度上值得参照。无论是最高院编选，还是外省高院编选，有参照价值的指导案例都应该遵循，二者的区别只在于位阶效力上。

4. 对于诉讼费用的监督规范

虽然诉讼费用有明确的计算标准，实务中的失误仍然在所难免，这种失误有时甚至达到非常离奇的程度。在一起 75 名员工与所在单位的股权转让纠纷中（代表人诉讼），由于立案庭错将该案的法律关系理解为劳动争议纠纷，争议标的额为 1900 多万万元的案件，最终只收取了 80 元诉讼费（本应收取约 10 万元左右）⁵。

当时未参加代表人诉讼登记的一名员工，之后于诉讼时效内起诉，争议标的额仅为 31.33 万，却被法院要求收取诉讼费用 6000 多元。对于同类案件（普通共同诉讼）的收费，后者争议标的额只是前者的 1.6%，但是诉讼费用反而多出 75 倍！这种失误明显超出了可容忍的范围。应用人工智能核算诉讼费用，能够避免因法官失误而造成的法院或当事人损失。

（三）给予自由心证必要约束与辅助

同时适用于两大法系的自由心证制度，其共性在于尊重法官的裁判权威，但是这种尊重不代表法官心证可以完全不受限制。否则，曾经举国哀痛的呼格吉勒图案件会反复发生（案件唯一“有力”的证据就是呼格指甲里的被害人血迹）。有学者将大陆法系自由心证制度的特点概括为“相对自由型”，将英美法系归纳为“严格约束型”。

重要区别之一是，“严格约束型”，法官对于证据证明力的权衡受到严格限制；“相对自由型”，法官则很少受到限制^[19]。日本现行《刑事诉讼法》中规定证据的证明力由法官自由判断。我国虽立法上未对自由心证作出系统规定，但张卫平教授认为：民事诉讼中证据证明力的判断交由法官自由心证没有争议，存在争议的是关于证据方法（证据资格、证据适格）的判断^[20]。

观察刑诉实践，基本与之相同。尽管从渊源、体例和方法上说，我国（社会主义法系或中华法系）更接近于大陆法系，但在自由心证制度的取向问题上，笔者更倾向于“严格约束型”。体现在人工智能评估裁判的应用上，可以在制度上作出设计：无论是权衡证据证明力还是判断证据方法，一旦发生人类法官判决与人工智能完全抵触的情况，应提交审委会讨论决定。如果是人命关天的死刑案件，甚至应该暂缓中院或高院判决的效力，由最高人民法院审理（防止地方利益集团勾结），而非简单的死刑复核。

人工智能还可以辅助法官自由心证。在证据认定方面，例如 2020 年的“腾讯诉老干妈案”，法官在案件裁判过程中未发现第三人冒用老干妈公章，以致作出了错误的财产保全裁定。在这类案件中，人工智能可以随时联网获取大数据，通过比较与鉴别公章在公安部门的备案信息，及时发现“萝卜章”问题，并在对法官的裁判评估中指出，帮助法官在自由心证方面弥补其能力之外的不足。

（四）复核难于计算的损失数额

法官不是会计师，却又要经常面对会计方面的问题。在民事诉讼中，并非所有的财产关系涉案金额都如合同关系般相对容易

计算。在侵权关系、劳动关系、商事关系中，涉及赔偿数额的计算相当专业，如股权置换金、养老金的计算，对于当事人提出的损失数额及其计算依据，在不具备会计专业基础的法官看来简直是无从下手，法官根据其“朴素的代数知识”判决之数额，当事人也往往不会认同。而涉及复杂会计方法的裁判，多集中在固定的某几种案件类型中。会计专家与法律专家完全可以开展合作，让会计专家也加入人工智能司法论证系统的开发中。通过跨学科的研讨，给人工智能系统建立一些常见的疑难损失额计算模型。

为防止法官因缺乏会计知识而错判，涉及复杂损失数额计算的案件，人工智能在评估法官裁判的同时，能够评估法官判决金额的合理程度。比如设定 20% 以内为合理误差，一旦出现人工智能评估损失数额与法官之判决超出 20%，则相差巨大，应当提请其他法官或审委会作出复核处置。

（五）快速推进司法质量提升

现行法官绩效考核制度存在诸多缺陷是毋庸置疑的，这些缺陷导致司法裁判质量某种程度的下降。虽然这一点早已形成共识，然而部分学者提议完全放弃考核转而依靠自律，认为法官职业应当高薪养廉^{[16]18-20}。不过，尽管考核方式不尽相同，但是几乎所有主流国家皆保留了这一制度。

不言而喻，针对法官断案绩效考核的最佳方法是，根据所审理案件本身裁判质量的好坏来评判。只不过在弱人工智能技术问世以前，存在三重现实障碍：第一，如果把一位法官的裁判拿来评估，就必须由一位评估人耗费差不多同样的精力将该案重新司法论证一遍。试想，在本就案多人少的状况下，谁来充当评估人？第二，假若另一位评估人评估的结果与法官大相径庭，以孰为准？第三，每位评估人的精力也极其有限，只能评估数量有限的案件。

那么，不同的评估人以不同的标准评估，所酿不公能服众否？而若以人工智能评估法官裁判，能够辅助法官更好地解决以上三个问题。首先，对比人类，人工智能进行司法论证的速度几乎可以忽略不计，而且不会额外占用法官这一紧缺司法资源。其次，考虑到国内法官办案条件所限（案件繁多，分配给每起案件的时间严重不足），难免忙里出错，人工智能进行司法论证通常正确率更高。最后，不言而喻，由人工智能评估案件裁判结果，标准统一、公平客观，大大提升了司法质量。

四、存在的法理难题及其解决方法

由于运用人工智能评估法官裁判（绩效考核）几乎未被现存文献讨论过，笔者也就只能预估其将会面临的质疑。

（一）并未突破法官独立审判的制度框架

与人工智能能否替代法官的质疑（机器可以审判人类吗？）类似，人们同样会提出疑问，机器可以评估人类吗？然而，这两者毕竟不同——法官裁判是司法过程，而考评法官裁判是行政过程。司法裁判权实质是一种独立判断权，不受外界干涉（包括上级）。如果将这一权力完全交由人工智能，人类将无法干预机器裁判，许多学者出于对人类社会失控风险的担忧是合理的。

而评估裁判、考核法官是行政权，不具有独立性，相反满足服从性。所以应用人工智能评估裁判、考核法官，正当性源于两点：一是评估程序由人类编写，评估方式与方法仍然受到人类意图支配。二是如果出现超出人类预期的评估结果，绩效考核作为行政权，不具有独立性，评估结果以及最终解释权则掌握在人类手中，绩效考核必须服从人类的复议和处置。

（二）使之不与法官自由裁量权冲突

另一方面的担心是，利用人工智能评估法官裁判，是否会导致法官在审判时蹑手蹑脚。法官很可能受制于被人工智能评估打分的顾虑，在其自由裁量权范围之内亦不敢作出顶格判决，而选择走中庸路线。这一点有点类似于高考作文阅卷，如果打出满分

或者零分，则会被单独提出接受阅卷组合议。

与其招惹麻烦和是非，不如给个大众分数安全。对于这种担忧，中国人民大学陈景辉教授的研究可以作为回应：“人工智能对于司法裁判的介入，基本上是通过建构模型的方式，将自由裁量或价值判断这个部分排除在外。”^{[1]41}

也就是说，大可不必担心一旦人工智能用作评估裁判质量，法官心理会受到逆向影响而在涉及自由裁量权时集体作出温和的判决（担心顶格判决受到人工智能的格外注意）。因为，从一开始，就可以在设计模型的时候使得人工智能在评估裁判时完全摒弃对触及法官自由裁量部分的评估。

（三）对于算法黑箱与算法偏见的把握

有观点认为智能司法的逻辑判断存在“黑箱化”，具有不可控因素^{[15]35-36}。本文并不认同。的确，人类无法还原人工智能的思维过程，不过无须大惊小怪。人类法官的思维过程同样也不可获知，但是能够知道表现于外的证明方法和分析过程。同理，我们只需清楚人工智能给出的证明过程，而不必在意其思维过程。

如此看来，用人工智能评估司法裁判仍然在现行制度框架内进行，具有制度连续性与安全可行性。值得一提的是，算法黑箱在某种意义上反映了人类的心灵黑箱，应警惕其与人工智能的算法黑箱所发生的相互作用。警惕以公正的初衷所做的努力实际上却带来新的、更为隐秘的支配力量，最终变为了不公正的结果。

人工智能的算法与人类的理性逻辑存在很大的不同，对于任何有学习能力的系统来说，它所输出的内容是由输入的数据所决定的，这并不是一个新的观点。比如，“罪犯再次犯罪概率预测系统”可能会得出评估结论——某省籍的人再次犯罪的概率大于其他省份的人。虽然整体上存在这种可能，但是对于个体而言却是明显的歧视。具体到法官绩效考核的应用领域，应当将其适用的场景细化，同时注重机器训练材料的选择，避免经验材料造成算法偏见。

（四）防止“数字控制”，维系“算法信任”

随着智慧司法系统的应用、中国裁判文书网的大数据公开，确实在某种程度上提高了审判效率和透明度，但公开并不意味着公信力的提升。于当事人个人而言，如果对判决结果不满意，并不会因为信息公开而信任裁判。

于社会公众而言，公开的裁判文书皆经过法官对于案件细节（包括案件事实、相关证据以及双方观点）的裁剪过滤，如若没有完整的卷宗，不可能判断法官裁判公正与否。朱景文教授指出，尽管一些官方数据表明司法机关所做出的成绩，但是人民群众如何评价这些成绩，与人们的日常感受是否吻合，仅有这些数据是不够的^[21]。而人民的满意度和司法公信力又是相得益彰、唇齿相依的。司法公信力使得当事人对裁判产生满意度，满意度又进一步提升了司法公信力。

司法公信力问题由来已久，人工智能可以破题。这源于现代生活中的“算法信任”观念。一个可以类比的例子是支付宝的花呗功能。这些技术对人们的信用度进行智能评估，全部交由算法来决定与处理。

尽管每个人不知道自己的信用额度是如何计算出的，也不知晓发生还款违约时罚息如何确定，但大多数人对于这种算法是信任的，尤其是年轻一代。同理，鉴于这种“算法信任”的观念，人工智能评估法官裁判的结果如若公开，比如裁判合格或者裁判良好，会对当事人的心理带来正面影响，将有助于提升司法裁判公信力，当事人甚至可能会因此放弃上诉。

遗憾的是，随着人类对于算法越来越多的反思，近年来人们对于算法的信任度急转直下。这期间，北京大学博士后陈龙关于“算法对于外卖骑手的劳动控制研究”一度引起了全社会的极大关注。他认为，由于外卖平台公司对控制权重新分配，使平台系

统与消费者取代了平台公司对骑手进行管理，平台公司退居幕后，看似放弃了对劳动的直接管理，实则淡化了雇主责任，还把劳资冲突转嫁到平台系统与消费者身上^[22]。

与之类似，法官绩效考核的人工智能系统使原本的人类考核部门退居幕后，看似放弃了考核管理权，实则同样是淡化了相关部门的责任，把冲突转移到人工智能考核系统与法官个人身上。因而，绩效考核的算法源码需要予以公开以及接受第三方专业机构的监督，才能够维系民众对于算法的普遍信任。

五、结论

本文主要探讨了运用人工智能技术评估司法裁判，替代现有法官绩效考核制度的可行性问题。我国司法现状不尽如人意，固然有起步晚的历史因素和人才少的现实困难，然而更多是绩效考核制度之固有缺陷导致。这是因为，人工智能应用于司法论证已经具备现实基础，这一技术真正困难的是如何把法律推理中已经存在的若干种知识前提精确地表述为机器语言。

不过，在如何教机器学会理解人类语言文本这方面，近年在科研上取得了突破性的进展。人工智能通过观察学习案例和不断演练裁判，不仅可以建立与人类相似的“司法经验”，而且积累的速度和数量级是人类不可与之匹敌的。

运用人工智能对法官进行绩效考核，能够摒除既有法官考核制度之弊病与解决既有法院审级制度之悖论，并且可以从程序监督、法律适用等方面规范诉讼过程。人工智能还能在约束自由心证，以及复核当事人难于计算的财产损失等方面发挥重要作用。

同时，大可不必担心使用人工智能考核法官会影响其自由裁量权。因为，从一开始在设计考核模型时，就可以让人工智能在评估裁判期间完全摒弃对触及法官自由裁量部分的评估。也不必过于担心“算法黑箱”引发的绩效考核不可控问题，我们只需清楚人工智能给出的证明过程，而不必在意其思维过程。最后，法官绩效考核的人工智能系统的算法源码需要公开或者接受第三方专业机构的监督。

参考文献：

- [1]陈景辉. 人工智能的法律挑战：应该从哪里开始?[J]. 比较法研究, 2018(5):136-148.
- [2]张保生. 人工智能法律系统：两个难题和一个悖论[J]. 上海师范大学学报(哲学社会科学版), 2018(6):25-41.
- [3]JORDAN BIGDA. The Legal Profession:From Humans to Robots[J]. 18 J. High Tech. L, 2018:396-409.
- [4]SEARLE J R. Minds, Brains and Programs[J]. The Behavioral and Brain Sciences, 1980, 3(3):417-424.
- [5]JACK COPELAND. Strong AI, Applied AI and CS[EB/OL]. [2018-12-2]. http://www.alanturing.net/turing_archive/pages/ReferenceArticles/what_is_AI/WhatisAI02.html
- [6]尼克·波斯特洛姆. 超级智能：路线图、危险性与应对策略[M]. 张体伟, 张玉青, 译. 北京：中信出版社, 2015:143.
- [7]SERGIO, DAVID, BECERRA. The Rise of Artificial Intelligence in the Legal Field:Where We Are and Where We Are Going[J]. Journal of Business Entrepreneurship&The Law, 2018(11):27-52.

[8]潘庸鲁. 人工智能介入司法领域路径分析[J]. 东方法学, 2018(3):109-118.

[9]纵博. 人工智能在刑事证据判断中的运用问题探析[J]. 法律科学(西北政法大學學報), 2019(1):61-69.

[10]RISSLAND E L. Artificial Intelligence and Law:Stepping Stones to A Model of Legal Reasoning[J].Yale L. J., 1990(99):1957-1981.

[11]BARNDEN J A,PETERSON D M. Artificial Intelligence, Mindreading, and Reasoning in Law[M]. Physica-Verlag H D, 2002.

[12]SMITH J C. Machine Intelligence and Legal Reasoning[J]. Chi.-Kent L. Rev, 1998(73):277-347.

[13]AI, 无师自通, 自学成才? [J]. 今日科技, 2017(10):16.

[14]李飞. 人工智能与司法的裁判及解释[J]. 法律科学(西北政法大學學報), 2018(5):32-41.

[15]钱大军. 司法人工智能的中国进程: 功能替代与结构强化[J]. 法学评论, 2018(5):138-152.

[16]李拥军, 傅爱竹. “规训”的司法与“被缚”的法官——对法官绩效考核制度困境与误区的深层解读[J]. 法律科学(西北政法大學學報), 2014(6):11-20.

[17]南京市中级人民法院课题组, 胡道才. 法官业绩考核评价制度研究[J]. 中国应用法学, 2018(1):121-135.

[18]李浩. 民事诉讼法学[M]. 北京: 法律出版社, 2016:308.

[19]相庆梅. 两大法系民事诉讼自由心证的司法适用及其启示[J]. 江西社会科学, 2017(3):188-195.

[20]张卫平. 自由心证原则的再认识: 制约与保障——以民事诉讼的事实认定为中心[J]. 政法论丛, 2017(4):14-21.

[21]朱景文. 司法满意度的社会评价——以 2015-2017 年法治评估数据为基础[J]. 中国应用法学, 2018(3):1-13.

[22]陈龙. “数字控制”下的劳动秩序——外卖骑手的劳动控制研究[J]. 社会学研究, 2020(6):113-135+244.

注释:

1 2017 年, 针对飞速发展的大数据、人工智能、区块链、物联网等科学技术, 中国人民大学法学院甚至专门成立了未来法治研究院。

2 这方面, 有观点认为, 虚拟货币、数据歧视、自动驾驶等智能技术、互联网新事物足以掀起一场“法律革命”。然而笔者并不同意这种观点, 诚然确实出现了一些新问题, 但大部分情形(如比特币、电子合约)却并未超出现有理论的框架, 虚拟财产仍然是财产, 电子合约依然适用合同法。

3 根据《人民法院报》2017 年 8 月 10 日第 002 版的评论, 智慧法院司法定位包含以下若干方面的功能: “其一, 诉讼流程

全在线。立案、证据提交、保全申请、开庭审理、电子签名、电子送达、执行申请等在内的网上一体化诉讼运行机制。其二，实现多平台信息共享与对接，工商、司法鉴定、公证机构等信息共享与对接，判决执行将与不动产登记、金融机构以及征信、资产处置平台等部门的系统进行对接，实现在线‘一键’启动开庭到裁判的执行。其三，智能化应用，充分利用算法以及司法大数据的优势，实现诉讼结果预判、类案推送、分析胜诉率，构建诉讼智能系统或者平台，引导当事人正确评估案件走向，体现专业化和智能化的审判体系。其四，诉讼诸要素模块化指引，复杂问题简单化、专业问题通俗化。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。”

4 中国留学生陈丹琦在斯坦福大学的博士学位论文《如何让机器学会理解人类语言》(Neural Reading Comprehension and Beyond)中，提出了如何解决机器学会理解人类语言的方案，引发空前反响，这也是人工智能最难以捉摸和长期存在的挑战之一。

5 参见“孔令民、张思侠等与淮北矿工总医院集团股权转让纠纷一审民事判决书”，(2016)皖 0603 民初 4763 号。