

司法人工智能的理论极限研究

陈敏光

(中国应用法学研究所, 北京 100062)

摘要: 司法人工智能有其理论极限。这种理论极限来源于三个方面。人工智能作为工具的不可能, 即不可能超越理性计算范畴、不可能脱离自身条件而发展、不可能超越既定历史而创新。人工智能与司法场景的不适, 也即无法适应司法的不可计算性、动态博弈性、能动创新性。人工智能对司法的型塑及其不可接受性, 其对司法形态的型塑包括诉讼从线下搬到线上、规则从物理走向网络、司法从解纷扩至治理, 要受到司法程序价值、司法本体价值的必要限制, 并有相关外溢风险的考量。文章立足理论极限, 并据此提出发展司法人工智能的若干建议。

关键词: 人工智能; 理论极限; 司法场域; 司法型塑

中图分类号: D922.1 **文献标识码:** A **文章编号:** 0257-0246 (2020) 11-0194-11

极限是数学当中的一种重要思想, 意指“无限靠近而永远不能到达”, 在数学分析和物理工程的应用较为广泛。^① 对于以数据概率论为基础的司法人工智能来说, 极限思想同样适用。只要在理论上厘定了司法人工智能的极限, 就能够更好地在边界里做事, 减少很多无谓的“妄想”和不必要的投入, 专注于机器智能优势的有效发挥, 并和人的主观能动性有机地结合起来, 实现人机互动和耦合的最佳状态。正是在这个意义上说, 厘定司法人工智能极限是确定应用之路的前提之一。目前, 这一理论极限被意识到了, 但未被清晰地整理总结。本文尝试厘定司法人工智能的理论极限, 主要包括三个维度: 客观上的不可能, 即人工智能作为工具本身的不能; 人工智能在司法场景中的不适性, 这是因为人工智能总是和特定的场景联系在一起; 主观上的不可接受, 即人工智能的司法型塑及其限度。当然, 客观上的不可能、场景上的不适性和主观上的不可接受在很大程度上交织在一起, 因为, 司法人工智能作为人工智能在司法场域中的应用, 是一个主客观相统一的过程, 既要体现工具的固有属性、潜能和作用, 也反映主体的价值追求或社会的可接受度。

一、问题的提出

目前关于人工智能的理解有两种截然不同的进路。一种进路是科幻意义颇为浓厚的“有机智能论”。一些未来学家、产业人士、媒体设想人工智能将经历弱人工智能、强人工智能和超人工智能三个发展阶段, 而在超人工智能阶段, 机器将具有自主意识, 从而威胁人的生存。如霍金担忧人工智能

基金项目: 司法部国家法治与法学理论研究项目 (18SFB3001)。

作者简介: 陈敏光, 中国应用法学研究所助理研究员, 博士, 研究方向: 法理学。

^① 参见夏征农、陈至立主编《辞海》, 上海: 上海辞书出版社, 第1019页。

的发展可能灭绝人类。^① 科技狂人埃隆·马斯克 (Elon Musk) 也坚信这一点, 并为此提出人机融合的方案, 并专门成立公司致力于将芯片植入人脑。^② 更戏剧性的是, 沙特甚至授予机器人以国籍。^③ 另一种进路则是“工具主义定位”。人工智能的本义是以机器的方式实现人脑的智能。^④ 姚海鹏、王露瑶等在《大数据与人工智能导论》中明确提出“从计算机应用系统的角度出发, 人工智能是研究如何制造智能机器或者智能系统来模拟人类智能活动的的能力, 以延伸人类智能的科学。”^⑤ 从实现的路径看, 人工智能的发展实践表明, “仿生学”的路径并不可行, 事实上, 本轮人工智能复兴的关键在于大数据、算法和算力特别是大数据的突破。^⑥ 故本文对人工智能做工具主义的定位和理解。

没有万能的工具, 且工具也在型塑人, 人工智能在司法中的运用 (也即司法人工智能) 自然也有其极限。^⑦ 从检索的文献来看, 人们对司法人工智能理论极限是有所意识的, 但在把握上较为粗糙, 尚缺乏系统的整理。学术界、实务界多热衷于司法人工智能的鼓吹、开发和应用, 这固然有国家政策导向的影响, 也有角色立场和思维方式的因素。如对司法人工智能开发企业来说, 鼓吹司法人工智能的应用前景并据此“攻城略地”占领市场, 就是其固有的商业思维所致, 从而在有意无意间忽略了“理论极限”的存在。又如按照一般的类推思维, 人工智能在其他商业领域的推广和取得的成功也会“冲动”地扩散到被动保守的司法场域, 从而掩盖了“理论极限”的存在。^⑧ 对上述问题, 学术界已有不少担忧和提醒。如马长山较为系统地论述了人工智能的社会风险和法律规制, 这些风险包括伦理风险、极化风险、异化风险、规制风险和责任风险等, 故需要塑造风险社会的法律理念, 建立多元互动的风险规制体系, 确立“过程—结果”的双重规制策略, 从而促进人工智能的健康发展, 构建智慧社会的法治秩序。^⑨ 王禄生则敏锐地捕捉到技术话语权与司法话语权的分离和技术权力对司法权力的介入限度问题,^⑩ 并呼吁司法大数据与人工智能开发须反思片面技术理性思潮,^⑪ 从而转向于司法属性对技术的规训。季卫东针对电脑量刑这一具体应用提出, 电脑法律实证主义的预

① Hannah Osborne, Stephen Hawking Warning: Artificial Intelligence could Destroy Civilization, <https://www.newsweek.com/stephen-hawking-artificial-intelligence-warning-destroy-civilization-703630>, accessed October 11, 2019.

② 埃隆·马斯克 (Elon Musk) 的逻辑是, 与其人类将来被人工智能消灭, 不如将人类和人工智能融为一体, 不分你我。为此, 马斯克专门成立了神经科学公司 Neuralink, 专门研究如何给大脑里植入芯片等脑机对接技术, 为人机融合做准备。Elon Musk launches Neuralink, A Venture to Merge the Human Brain with AI, By Nick Statt, <https://www.theverge.com/2017/3/27/15077864/elon-musk-neuralink-brain-computer-interface-ai-cyborgs>, accessed October 13, 2019.

③ 2017年10月, 在沙特阿拉伯利雅得举行的一次以创新为主题的专业会议上, 该国宣称自己是世界上第一个赋予机器人公民身份的国家, 引发关注。参见《首位人类国籍机器人“索菲娅”获沙特公民身份》, http://www.cssn.cn/gj/gj_gjzl/gj_ggzl/201710/120171031_3689730.shtml, 2018年10月12日访问。

④ 人工智能这一概念, 系数学家约翰·麦卡锡、计算机与认知科学家马文·明斯基、IBM 系统设计师纳萨尼尔·罗切斯特和信息论创始人克劳德·香农等在举办“达特茅斯夏季人工智能暑期研究会议”时提出, 当时的会议策划书表明“本次研讨会将以这样一个猜想为基础而展开: 学习的每一环节以及智力的其他方面原则上都可以得到如此准确的描述, 以至于我们可以制造出一台机器对其进行模拟。我们将尝试发现如何让机器使用语言、提出抽象命题和概念、解决某些目前留给人类去解决的难题以及实现自我改进……” John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf>, accessed October 21, 2018.

⑤ 姚海鹏等《大数据与人工智能导论》, 北京: 人民邮电出版社, 2017年, 第4页。

⑥ 吴军将人工智能的定义分为狭义和广义两种。狭义的人工智能专指20世纪五六十年代特定的研究机器智能的方法 (即人工智能1.0版), 坚持机器人要像人一样思考才能获得智能, 并形象地称之为“鸟飞派”。而广义的人工智能还包括通过数据驱动、知识发现或者机器学习等现代方法来实现的机器智能 (不妨称之为现代机器智能), 这是工程学意义上的人工智能。在吴军看来, “鸟飞派”正如人通过模拟鸟的飞行而制造飞机, 而不是利用科学的空气动力学原理, 因而被实践证明并不可行, 而以数据驱动为基础的机器智能是目前真正的人工智能。参见吴军《智能时代》, 北京: 中信出版集团, 2016年。

⑦ 正如马歇尔·麦克卢汉所言“我们自身变成我们观察的东西……我们塑造了工具, 此后工具又塑造了我们。”参见马歇尔·麦克卢汉《理解媒介——论人的延伸》, 周宪、许钧译, 北京: 商务印书馆, 2000年, 第2页。

⑧ 笔者曾于2018年4—6月间参与某业界知名公司开发司法人工智能产品的前期论证工作, 以上切身体会即源于此。

⑨ 马长山《人工智能的社会风险及其法律规制》, 《法律科学》2018年第6期。

⑩ 王禄生《大数据与人工智能司法应用的话语冲突及其理论解读》, 《法学论坛》2018年第5期。

⑪ 王禄生《司法大数据与人工智能开发需反思片面技术理性思潮》, 《民主与法制时报》2018年6月3日, 第6版。

设不恰当地排除了诸如自然法、权利保障、天理人情,又较少注重预防之类的思辨性要素,忽视了司法运行当中所必须考量的背景条件和地方性知识等。在某种意义上说,上述思辨性要素、背景条件和地方性知识就标示出计算机法律专家系统软件以及人工智能审判的疆界或者局限之所在。^①

应当指出,一方面,上述研究主要还是抽象意义的,并未对司法的属性进行系统反思和展开,从而也就无法充分地探讨司法应当如何规训技术。另一方面,司法人工智能工具属性背后的技术原理应当进一步呈现,从而有利于人们发现其本身的工具极限和与司法之间的不适性等。故恐怕还要立足既有研究,对司法人工智能的理论极限做更加系统而深入的探讨和厘定。实际上,我们需要回答一系列环环相扣的问题:司法人工智能因其工具性限制不可能为司法做哪些事?在司法人工智能能够做的事情中,需要对司法改革、司法制度、司法审判、法院管理、社会服务等做何改造,成本、效益、风险等是否可知、可靠、可控?经由技术改造后的司法是不是我们需要和期待的司法?不难发现,这一系列问题实际上确定了司法人工智能理论极限的三个方面:人工智能在工具意义上的不可能(不能),人工智能在司法场域中的不适(不适),司法对人工智能型塑的不可接受(不该)。下文就这三个方面详细展开论述。

二、人工智能作为工具的三个不可能

尽管和以往工具相比,人工智能正在迭代发展,展现出相当程度的自我适应性,但从技术原理来看,其并非万能,仍存在着不可逾越的极限。也就是说人工智能在工具意义上存在极限,具体说,人工智能在技术上存在三个方面的不可能,也即不可能超越理性计算范畴、不可能脱离自身条件而发展、不可能超越既定历史而创新。

1. 人工智能不可能超越理性计算范畴

上文对人工智能做了工具主义的定位。这就意味着,司法人工智能不可能在主体性上完全替代法官和其他司法工作人员,其所能发挥作用的领域主要是理性计算范畴。就人的主体性的其他方面而言,如感性、意志和信仰等,人工智能并无所长。就感性层面而言,尽管在外观上看,人工智能似乎具备了一定的感知技术能力,如语音识别技术对应于人的听觉,图像识别技术对应于人的视觉,自然语言处理技术对应于人的读写,但无论从数据科学还是智能科学而言,这些技术都属于“数据搜集”阶段,即将格式多样的数据收集、汇聚,继而服务于后续的“认知”阶段。^②通俗地说,人工智能只是运用数据搜集、分析和计算的机器方法,实现了人脑的智能而已,它本身并不理解人脑的运作机理。希尔勒(John Searle)提出“中文房间”的假设,很好地诠释了这一点:房外的人测试人工智能系统是否真正理解其手中的象形文字,而身处房间的人工智能系统,只是根据象形文字的对照手册或字典机械地告诉房外人具体答案。虽然它给人以智能的印象,但其无法真正理解接收到的信息,或者说对象形文字一窍不通。^③至于意志、信仰等,人类尚且发出“我们要悬置知识,为信仰留下空间”的感慨,司法人工智能恐怕更无法企及了。

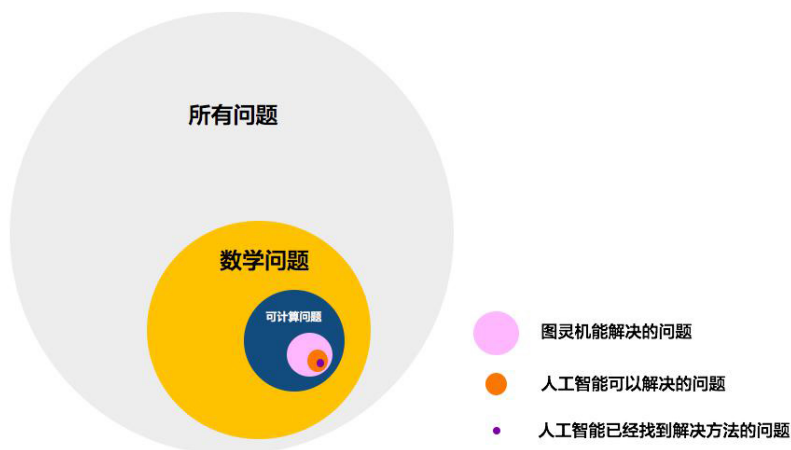
就理性计算范畴而言,人工智能也不是万能的。要理解这一点,不妨上溯到计算机的本源。计算机之父图灵设计了“图灵机”这一数学模型,为今天和未来很长时间的计算机划定了一条不可超越的界限,遵循以下底层逻辑原理:世界上是否所有数学问题都有明确的答案?如果有,是否可以通过有限步骤的计算得到答案?对于那些有可能在有限步骤计算出来的数学问题,能否有一种假想的机械,让它不断运动,最后当机器停下来时,那个数学问题就解决了?图灵最终研究的结论是,世

^① 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 华宇元典法律智能研究院编《让法律人读懂人工智能》,北京:法律出版社,2019年,第13-14页。

^③ 华宇元典法律智能研究院编《让法律人读懂人工智能》,北京:法律出版社,2019年,第13-14页。

界上有很多问题，其中只有一部分是数学问题；在数学问题中，只有一小部分是可解的；在可解的问题中，只有一部分是理想状态的图灵机可以解决的；在后一类问题中，又只有一部分是今天实际的计算机可以解决的；而人工智能可以解决的问题，又只是计算机可以解决问题的一部分。^①



人工智能解决问题的相关图示

可见，从本源设计上看，作为计算机应用系统的人工智能所处理的无非可以解决的部分有解的数学问题，而所根据的技术原理无非大数据的收集、分析和发掘等，只不过它以不同于人脑的这种方式实现了人脑的智能而已。故目前的人工智能技术不可能超越理性计算范畴。

2. 人工智能不可能脱离自身条件而发展

与传统工具相比，人工智能并非固定不变的工具，而是具有相当程度的自我适应性或说自我改进性。这种自我改进性来自于机器学习算法，尤其立足于神经网络技术而发展出来的深度机器学习。与传统线性建模算法相比，此类算法完全利用输入的数据自行模拟和构建相应的模型结构。这一算法决定了它是更为灵活地依据不同的输入来训练数据而拥有的自优化特性。^②正是这种自我改进的智能发展，让人有种杞人忧天式的担忧——人工智能的学习能力将越来越强，从逻辑上推演，它终究超越甚至取代人类。然而，这恰恰是一种没有极限思维的单线想法，如果理解了上述计算机的本源设计和技术原理，就很容易看出其中的谬误。更具体地说，人工智能的发展并非如庄子所言的“无所待而逍遥游”，而是要受制于大数据、算法、算力三要素等自身条件的限制。

就大数据而言，正如前文所提到的，本轮人工智能正是因为遇到大数据才得以迅猛发展。尽管大数据具有海量（Volume）、高速（Velocity）、多样（Variety）、真实（Veracity）等特征^③，但这也是理想状态或极限状态的描述。现实中，数据的采集总会有盲点或误区，因其与主体的“先验认知”密切相关，而采集与人相关的数据就更复杂了，它们自然不如物理之物那样稳定、明确。在某种意义上说，采集数据并不是“摘果子”那样简单，在其后隐含着主体视野、价值取向和社会关系的可信性等复杂命题。

就算力而言，其是算法有无“用武之地”的重要外在保障。姚海鹏等认为，作为深度机器学习基础的多层神经网络在计算机运算能力取得突破之前，几乎没有实际应用的价值，因为运算量太大了。在十几年前，用神经网络算法计算一组并不海量的数据，辛苦等待几天都不一定会有结果。^④在

① 吴军 《谷歌方法论》，来源于“得到 APP”。文中的图示亦来源于此。

② 姚海鹏等 《大数据与人工智能导论》，北京：人民邮电出版社，2017年，第6页。

③ 蒋惠岭 《司法大数据能给我们带来什么》，《中国审判》2018年第3期。

④ 姚海鹏等 《大数据与人工智能导论》，北京：人民邮电出版社，2017年，第6页。

今天,得益于摩尔定律和与此相关的硬件设施的发展,算力得以大大提升,从而在很大程度上突破了人工智能的发展瓶颈。

就算法而言,要依靠前端训练集的取样分析,并通过后端的模型形成和测试,根据新的数据如此循环反复,才能得以不断优化。然而,这也是理想状态。实践中,常常会存在模型的过度拟合——只在训练样本上表现得很好而不能对于新的数据样本做出很好的预测或模型的欠拟合,由于数据集太小导致模型不能预测整个样本空间。此外,在机器学习当中还存在 no free lunch 定理——即不存在能够在所有可能的分类问题中性能均为最优的算法。例如序列 1, 2, 4, 既可能是“前数+1, 前数+2, 前数+3”的规律,也可能是“2 的 0 次方, 2 的 1 次方, 2 的 2 次方”的规律,而无论何种算法,显然无法同时满足两种规律。^①

3. 人工智能技术不可能超越既定历史而创新

尽管人工智能的目的并不在于对既有训练样本做出更好的解释,而是对未知样本做出更好的预测。且人工智能的工具属性确有优势,这主要表现为信息的多维度,即通过多维度的信息汇聚分析,帮助人们提高认知高度,尽可能地消除认知盲点、认知误区或认知偏差。考察发现,人工智能对未知样本的“泛化”能力在本质上是一种归纳推理,也即由有限的数据集(特例)推广到整个样本空间(一般)。^②而从大数据角度看,人工智能所处理的是既有或正在实时生成的数据,而非尚未生成的数据。也就是说,它所依据的是人类既有的经验——即人们常说的“有多少人工就有多少智能”或较为负面的说法“垃圾进垃圾出”,故人工智能技术不可能超越既定历史而创新。这实际上也是人工智能不能完全代替人类主体性的重要方面。与人工智能相比,人类能够感知具体经济社会条件的变化,而非对其做程序化、格式化的机械理解,从而能在一定条件下摆脱既定行为模式,创造性地推动历史向前发展。

综上,可以看到人工智能自身局限性和基于此的应用限制。正如中国人工智能领域泰斗张钹院士所总结的“目前的人工智能技术适合解决的问题,一般都要有充分的数据和知识储备,而且需要的是确定性问题、完全信息、静态的、有限任务和特定领域。”^③

三、人工智能在司法场域中的三重不适

人工智能的应用总是和特定的场景联系在一起。如果将人工智能的三种不可能与司法场域的特定属性结合起来,就可以清晰地发现其在司法场域中的三重不适性,这主要包括无法适应司法的不可计算性、无法适应司法的动态博弈性、无法适应司法的能动创新性。

1. 无法适应司法的不可计算性

人工智能不能超越理性计算范畴,而司法并不仅仅是或主要不是理性计算活动,它是人与人之间的直接社会实践活动,故在应用上,两者存在难以跨越的鸿沟。

从实证分析法学的角度看,司法具有规范性、程序性和逻辑性,它要求法官以事实为依据、以法律为准绳进行裁判,不得“擅动界碑”,且裁判的过程应符合基本的程序要求,满足逻辑一致性的要求。从这个意义看,司法确实在某种程度上具备“可计算”的基础。然而,并不存在“纯儿又纯”的法律规范,在这些规范背后站着主体的主观愿望、价值诉求和政策考量。^④这就意味着,司法裁判不是程序化的理性计算,而是事实与价值的复合体(实体上)和技术应用与民主过程(程序上)的统

^① 姚海鹏等《大数据与人工智能导论》,北京:人民邮电出版社,2017年,第26-30页。

^② 姚海鹏等《大数据与人工智能导论》,北京:人民邮电出版社,2017年,第29页。

^③ 华宇元典法律人工智能研究院编《让法律人读懂人工智能》,北京:法律出版社,2019年,第8页。

^④ 史际春《由民法看法本位》,《法学家》1992年第1期。

一体，它要借由事实查明和法律适用这一过程体现人性温度、彰显人文关怀。从实体法角度看，对犯罪与否的定性、刑罚的科处、直接涉及伦理道德的家事案件的办理，都要首先考虑社会效果问题，再用法律规范的技术主义予以证成，实现法律效果和社会效果的统一。即使是那些纯粹计算的司法认定，也不能不受某种司法立场或价值判断的影响。这些变量恐非程序化的人工智能所能具体吸纳并加以有效调适。从程序法角度看，看得见的正义也同样重要。这不仅是因为司法本身需要过程来完成事实认定和法律适用，更是因为司法民主的内在需要，司法的中立性、被动性、公开性、亲历性、对抗性、集中性、终结性等也只能依靠具有感情感知、价值体认的法官去实现。

上文阐释了司法的不可计算性，做一个不甚恰当的类比，正如牛顿所感慨的“我可以计算天体运行的规律，却无法计算人性的疯狂。”让不具备情感、意志的人工智能来承担这一职能，更是勉为其难。“目前和将来相当长时期的人工智能仅仅是辅助性司法活动的实践工具、司法裁判活动的新分析工具而已，其本身尚无法感知、承载和落实司法根本属性所凝结的法的伦理道德诉求、政策价值考量等。”^①

2. 无法适应司法的动态博弈性

人工智能技术受到大数据、算法、算力等自身条件的限制，而司法的动态博弈性对上述各方面提出了更高的要求，致使司法人工智能的应用有难以克服的瓶颈。

从实际运行角度看，司法并非静态封闭的，而是动态博弈的一个过程。具体地说，法官居间裁判，控辩双方则不断地搜集证据、举证质证，罗列和阐释法律上的根据，试图说服法官并做出对自己有利的裁判。从信息角度来看，随着控辩裁三方的不断博弈，越来越多的和案件相关信息将不断汇聚。从理论上讲，这将越来越逼近“案件真相”，也完全符合利用信息的认识论原理——用更多更丰富的信息以消除信息的不对称性。^②但问题在于，控辩双方很可能基于自身的立场、利益整合出两套逻辑自洽的论证体系，并选择性地提供信息、隐匿相关案件信息或增加若干干扰信息，且随着审判进程的发展，各种不同价值或不同干扰度的信息都会继续“膨胀”，如何对不断输入的信息进行有效判断，并在恰当的时点“关闭”信息的继续“流入”，就是法官的智慧和经验所在了。可以看到，司法领域的信息状态明显不同于人工智能所擅长的“围棋领域”。棋盘已经划定了可能的步数范围，棋手对弈的信息也完全受围棋规则限制并展现在一步步的“落子”当中，也就是说不存在选择性的不完整信息或干扰信息，故人工智能只要根据这些充分封闭的信息计算出“最优落子可能”即可。

将围棋领域的人工智能简单地平移适用到司法场域，就可能会导致“橘生淮南则为橘、橘生淮北则为枳”的后果，根据片面或错误的信息得出片面或错误的结论，而无论其根据的大数据如何庞大和丰富、算法如何有效科学、算力如何快速有力等。表面看来，关键的是输入信息的分析和判断，但更深层次的是对人性、具体社会关系、生活经验的体察和把握，而只有人才能更好地理解人，作为工具的司法人工智能应止步于此。

3. 无法适应司法的能动创新性

从性质上看，人工智能是对既有经验或数据的提炼和升华，故其不可能超越既定历史而创新，这恰恰与司法的能动创新性不相适应。相较于立法的一般性、行政的主动性，司法通常表现为个案性和被动性，在运行当中又因其旨在为权利提供救济而必须保守中立性。故从总体而言，司法是保守的，

^① 冯洁 《人工智能对司法裁判理论的挑战：回应及限度》，《华东政法大学学报》2018年第2期。

^② 熵是和信息对立的一个概念，是对系统混乱程度（不定度）的一种度量，而信息是对消除这个系统的混乱程度（不定度）的能力的一种度量。因此，熵标志着系统的无序状态，而信息标志着系统的有序状态……至于有序和无序，在事实上是根本无法简单地用信息量的多少来说明的。这仅仅是有用信息和无用信息之比在人类认识中的某种反映，所谓有序就是有用信息占比例多些，所谓无序就是有用信息占比例少些。参见邬焜《信息哲学——理论、体系、方法》，北京：商务印书馆，2005年，第506-510页。

而司法克制主义也应成为法官的主流意识形态。^①但在现代经济社会条件下,这种主流意识形态受到了不同程度的冲击。一方面,伴随着社会关系网络化、立体化的发展,狄骥所言的主体之间的“社会连带”性日益紧密,公权与私权呈现出融合发展的趋势,社会对司法的期待也不限于个案纠纷的解决,而是要在个案处理中更好地回应社会,并尽可能地在社会治理中发挥更大的作用。另一方面,一些新型的社会关系随着平台经济的发展而催生,既有的传统法律关系均难以对其完全适用(如专车平台与司机端的法律关系认定^②、数据竞争的司法裁判问题等),在立法滞后和行政执法不足的情况下,如何巧妙地运用司法政策、法律原则精当地处理相关案件,并在实质意义下推动制度供给,实现制度对社会经济发展的助推,就成为摆在法院面前的重大课题。这就意味着,司法在某些特定情况下要秉持一定程度的司法能动主义,以“熨平”社会关系与法律之间的脱节,而相应的裁判思维也不应局限于机械的法条主义,而是要有更宽广的社会、经济维度的考量,并智慧地将此种考量和法律原则、司法政策、相关法律规范等有机地融合在一起,从而满足法的自治性和安定性。这种努力代表了司法的能动创新性,而以既有经验或数据为基础的司法人工智能,注定只能重复本质相同的原理,无法提供面向未来的具体方案。

四、人工智能的司法型塑及其限度

人工智能技术嵌入司法制度和其运行当中,潜移默化地型塑着司法的基本形态,这既有积极的一面,也有消极的一面,同时还伴随着技术迭代发展所带来的不确定性或技术失控所带来的风险。在某种意义上,那些与司法不契合或司法无法接受的改变就构成司法上的极限,这实际上也是司法对技术的规训。

1. 司法型塑的基本方面

如果将司法视为一种“社会生产力”的话,人工智能技术就是新的“司法劳动工具”了,而新工具的引入必然带来司法基本形态的改变。

(1) 诉讼从线下搬到线上。司法人工智能的发展是以前期的司法信息化,即电子化、网格化为前提的。具体地说,人民法院信息化1.0版以“数字化”为核心,关注纸质文档向电子文档的转化,人民法院信息化2.0版则以“网格化”为核心,关注法院工作方式由线下向线上的转化^③,两者共同为司法人工智能奠定了硬件和软件基础,实现了法院系统层面的互联互通。人民法院信息化3.0版则以“智能化”为核心,要考虑的重点是,如何更有效地利用所留存和正在生成的各类数据,通过智能化的处理,更好地应用于司法审判、法院管理和社会服务。因而,人工智能对司法基本形态的型塑也应遵循信息化改造的基本逻辑,将诉讼从线下搬到线上,具体包括网上立案、电子送达、在线审判和审判管理的信息化、流程化等。而线上诉讼活动又在实时地生成和积累着大量数据,并不断地驱动着相关司法人工智能的应用,如自动审查起诉材料是否齐备完整、依靠语音识别技术自动进行庭审记录、人工智能辅助科学证据可采性审查、裁判文书样式自动生成、类案类判推送、辅助评估保释、预测案件审理结果等。循此逻辑,最高人民法院在加强智慧法院建设中提出,要努力实现法院“全业务网上办理、全流程依法公开、全方位智能服务”。

(2) 规则从物理走向网络。司法人工智能无法离开网络空间而运行。基于网络空间的相对独立性,传统的诉讼法规和诉讼流程、实体法规则乃至法院组织形态都在不同程度上受到了技术上的冲击。这种冲击在诉讼规则和诉讼流程上体现得最为明显。如电商买卖交易纠纷中的管辖权规则问

^① 陈金钊认为,司法克制主义最适宜法治理念的实现,因为只有司法奉行克制主义才能保证法律意义的安全和稳定。参见陈金钊《法官司法的克制主义姿态及其范围》,载《法律方法》第7卷,济南:山东人民出版社,2008年,第35-44页。

^② 陈敏光《平台经济背景下劳动关系的变革——以专车业态为例》,《现代管理科学》2018年第9期。

^③ 王禄生《司法大数据与人工智能技术应用的风险及伦理规制》,《法商研究》2019年第2期。

题,无论依原告就被告的一般地域管辖抑或依签约地、履行地、被告住所地的特殊地域管辖,都无法从根本上解决社会成本过高的问题,而根据技术对技术的思维,“在线管辖权”及相应配套制度的构建就能很好地解决这一难题。不止于此,具有互联网特性的案件的证据和当事人之间的沟通等,也主要产生和存储于互联网,由此产生了新的问题——电子证据的鉴真、审查、保全、认证和审理方式变革的必要性等。对这些问题,现行法律和司法实践均有一定程度的回应,如浙江、北京、广州互联网法院的增设和《最高人民法院出台互联网法院审理案件规定》的出台,大体确定了网上案件网上审理的思路,^①认为这既方便诉讼,又有助于通过审判创制依法治网规则。在电子证据方面,修改后的三大诉讼法均将其作为新的证据形式,后续亦有相关司法解释的陆续出台,如2014年最高人民法院、最高人民检察院、公安部《关于办理网络犯罪案件适用刑事诉讼程序若干问题的意见》(简称“意见”),2016年9月由最高人民法院、最高人民检察院、公安部联合颁发的《关于办理刑事案件收集提取和审查判断电子数据若干问题的规定》(简称“规定”)等,进一步充实了电子证据的相关规则。

在实体法上,这种冲击亦不难寻觅。近年来关于网络法研究的兴起正是例证。网络法其实并非传统意义的部门法,而是一个新兴的领域,传统部门法都不同程度地受其冲击,如民事领域中网络交易平台经营者的民事责任问题、网络效应与反垄断法的执行和刑事领域中共犯理论的可适用性等。Lawrence Lessig 提出,网络世界中也同样存在四种规制手段,分别是法律、社会规范、市场、代码,而代码在一定情况下可以代替法律。^②上述问题都引起了实体法规则来源和具体内容的变化。

至于法院组织形态,现有的互联网法院的发展已经体现相应的变化,正如有学者所指出的“杭州互联网法院的设立是司法机关适应信息化时代变革、积极承接全面依法治国战略和网络强国战略司法责任的结果,同时也是司法积极对接信息时代挑战的结果,适应了网络时代的诉讼特点,进一步推动了司法效率与公平。”^③而深入审判组织和具体司法运用层面,我们可以发现更为隐蔽的变化。例如,随着“要素审判”等司法人工智能系统的研发、利用,算法决策将在事实上“分享”法官的司法裁决权,正如上文所论述的,它也将在很大程度上冲击着司法的被动性、公开性、透明性、多方参与性、亲历性、集中性和终结性特点。

此外,如何在合议庭层面摆正法官、法官助理、书记员和司法人工智能系统的关系,进而如何调适合议庭、审判委员会和上下级法院之间的关系,也将是面临的难题。季卫东在中国法学会审判理论研究会2017年年会暨“两法”修改会议上指出,人工智能时代的变化并没有及时和充分地反映到《人民法院组织法》和《法官法》的修改当中,具体包括数字信息处理部门、电脑工程师等司法辅助部门及其人员的权利义务、法律和计算机程序编码以及大数据算法的支配下如何强调法官“有思考的服从”,判决自动生成的技术如何与法律解释和法律沟通兼容等。^④

(3) 司法从解纷扩至治理。人工智能将超越传统司法的定纷止争功能,更好地拓展了司法的社会治理功能。从系统论看,司法不仅仅是个案的裁判,亦是社会治理体系的重要一环。通过个案裁判形成制度助推往往是拓展司法社会治理功能的重要手段,由于各国司法制度、法律传统的差异,这往往呈现出不同的表现形式:判例法国家的判例本身就蕴含着具有法律效力的规则体系,司法直接介入社会治理。非判例法国家的司法裁判的介入则可能要间接和迂回一些,如中国法院系统更可能通过司法建议、调研报告、新闻发布的方式介入其中。当然,必须指出,司法解释是例外。在大数据和人工

^① 《最高人民法院出台互联网法院审理案件规定》第1条规定,互联网法院采取在线方式审理案件,案件的受理、送达、调解、证据交换、庭前准备、庭审、宣判等诉讼环节一般应当在线上完成。根据当事人申请或者案件审理需要,互联网法院可以决定在线下完成部分诉讼环节。

^② “Lawrence Lessig, ‘The Law of the Horse: What Cyberlaw Might Teach,’ *Harv. L. Rev.*, Vol. 113, No. 50, 1999.”

^③ 于志刚、李怀胜《杭州互联网法院的历史意义、司法责任与时代使命》,《比较法研究》2018年第3期。

^④ 景玥《法学大咖共聚金陵论司改成果入法如何除旧布新》,人民网, <http://legal.people.com.cn/n1/2017/0910/c42510-29525963.html>, 2019年4月12日访问。

智能背景下,司法的社会治理功能得到了很好的拓展,尽管这方面的认识和研究较少,但国内已有法院通过人工智能技术形成社会纠纷根源的数据研究报告,并积极参与府院相协调的源头治理。这种源头治理很好地把握和运用了司法人工智能信息多维度的工具属性,摆脱了“案海战术”的困境,从社会治理层面最大限度地提升了司法效率。根据笔者对北京高院的调研,主管副院长和有的实务人士就认为:北京地区民事法官年人均结案400余件,其中相当一部分是供暖合同纠纷、物业服务合同纠纷等法律关系相对简单的案件,不堪重负。而通过大数据和人工智能的技术分析,此类纠纷集中在特定的区域,与物业管理、房屋产权状况等密切相关。如能据此进一步分析成因,推动政府、法院参与协同治理和源头治理,则可以从根源上减轻法官工作负担,将其从“案海战术”中解放出来,腾出更多更宝贵的司法资源处理其他更为重要的纠纷。正是基于这一点,最高人民法院于2016年11月10日研究组建了“中国司法大数据研究院”,专设司法大数据分析中心,并定期向最高人民法院呈送相关分析报告,进一步充实和拓展司法的社会治理职能。

2. 司法型塑的必要限度

人工智能介入司法内部,作为一种潜在和内生的力量型塑着司法,将(程序意义的)诉讼从线下搬到线上、(实体意义的)规则从物理走向网络、(功能意义的)司法从解纷扩至治理,贯彻着技术主义的逻辑,在各个层面上体现出强大的效率导向。这在一定程度上是以牺牲司法的价值和属性为代价的,因而必须将之控制在必要的限度之内。

(1) 基于程序价值的限制。正义不仅应当实现,还应当以看得见的方式实现。司法意义的程序绝不同于计算机意义的程序,它正是要在一步都不能少的步骤中充分体现出司法的民主价值,让参与其中的当事人和社会公众感知到司法的中立性、平等性、专业性、权威性和便利性等。而这种感知又将反过来进一步强化司法威信。很大程度上,司法既不掌财,也不掌兵,其权威的社会认同恰恰在很大程度上来自程序公正的感知和体验。恰如有学者认为的“法院审判的独立性是基于政治权力对法治原则的认可而存在,尽管法院自身既无财权也无军权;司法的权威性除了基于自身的强制力外,更是依靠广大人民群众的信心和信任;司法的终局性则主要基于法律的程序正义原则,把对事实和法律问题的判断止于程序救济终结之处。”^①而如果司法一切都如计算机那样被程序化为材料或指令的输入或输出,就无需当事人的参与,甚至司法或法官也没有存在的必要,或者准确地说,司法已异化为计算机及其运行。将诉讼从线下搬到线上,最为显著的变化就是司法场景的转化,进而导致以下几个方面的程序价值的减损:由法台、国徽等法庭设施所营造的“司法剧场效应”受到削弱,法庭的威严感、神圣感难以感知。既有诉讼程序中的亲历性、对抗性、规范性、严密性等因网络的非现场性而不可避免地受到了削弱,并进而产生一系列相关问题,如实践中的异步审理是否有违诉讼法的直接言词原则、证人在线作证的有效性和证据出示的有效感知和质证等问题。司法程序价值理应成为司法人工智能应用的界限,除非此类问题有更好的替代解决方案,或对其容忍符合司法效率和程序价值的权衡取舍。

(2) 基于司法本体的限制。从本体上看,司法是指运用国家权力对诉讼案件进行审判的活动。司法即国家的审判活动,司法权就是审判权(也即一种居中的裁判权),通过个案裁判的方式维护法的价值的终局性的权力,而司法机关就是审判机关。^②然而,随着大数据和人工智能技术的嵌入,不仅外在的诉讼活动、程序和规则发生了变化,内在的司法主机(司法主体)、司法权也在悄然改变。可以预见,在当前法官普遍受制于技术门槛而不能很好地规训技术滥用和失控的情况下,技术人员有可能通过算法的构建和引入,在事实上“分享”司法权,从而在一定意义上成为“影子法官”。更复杂的是,这种算法在技术上存在着不可解释性(算法黑箱),在价值上又不可避免地带有技术人

^① 蒋惠岭《司法改革进入新时代 六大配套工程应当成为改革重点》,《人民法治》2018年第1期。

^② 周玉华主编《中国司法学》,北京:法律出版社,2015年,第16页。

员的前见甚至是偏见（算法歧视），并以机器的规模化方式^①驱动司法之运转，这恐怕不是现行司法制度能够接受的，也不为社会公众所认同。而此种隐患已经开始显现或可以预见。例如，威斯康星州发生的 *State v. Loomis* 案中，COMPAS 量刑系统被认为存在算法上的种族歧视。^② 即使不存在上述问题，技术人员在司法自身逻辑框架内所开发的“要素式审判”“知识图谱”等技术应用，也会因认知的局限而存在“简化实体审判”的问题，如道路交通事故损害赔偿纠纷中精神损害抚慰金的确定就不应当是要式化的逻辑推理，而是要顾及侵权人恶性、受害人受损等具体情境加以通盘考量。

进一步地说，基于司法本体的限制要回答的是以下几个层次的问题：司法改革对人工智能技术的强化是否会导致司法权本身被分化为司法权力和技术权力？如何在法院组织层面对司法权进行有效配置？人工智能技术人员是否将成为新一类司法辅助人员？其与法官的关系又当如何定位？而对这些问题的回答就构成了司法人工智能另一界限。

（3）基于外溢风险的限制。基于信息的多维度，司法人工智能确实能够大大拓展司法的社会治理功能，从而赋予定纷止争功能以更广阔的社会意义。司法大数据不同于商业大数据，具有公共属性，数据的公开、存储、流转、开发、利用应以何种主体为依托、以何种方式进行，均是有待深入探索和思考的重要问题。商业领域的数据垄断、数据杀熟会以同样的原理和方式表现在司法领域，如一旦智慧法院建设当中的平台设施和相应的数据要受制或依赖于网络技术公司，则当事人信息（包括身份信息）甚至审判秘密的安全等也可能受到威胁。侯猛就此认为，利用互联网技术进行网上审判的创意，最早就来自于阿里巴巴公司，且杭州互联网法院是依托阿里巴巴建立诉讼平台来运行的，这既影响了阿里巴巴等平台公司与其他公司之间的非对等地位，如阿里巴巴公司的竞争者为防止数据被监控或泄露，会拒绝在该法院进行网上审判，也增加了法院自身被技术公司俘获的可能。^③ 另外，司法大数据是以司法公开为条件的，其本身并不涉及正当与否的问题，关键在于开发利用主体及其目的。作为与国家政权勾连的权力的一种，司法不可避免地带有政治的“烙印”及相应的意识形态色彩，这就决定了对司法大数据的智能运用应有目的上的正当性，尤其国家安全的考虑。利用大数据等智能分析的手法，用诸策划“颜色革命”或开展其他政治渗透等活动，早已不是新闻。在民族国家的背景下，政治斗争、经济竞争必然长期存在，别国利用可以随手获得的司法大数据，按照其目的，经由人工智能的处理，进行相应的政治渗透、意识形态层面的“造势”，也并非不可能。且从现实来看，中国目前尚停留在技术应用层面，与此相关且更为重要的网络基础架构、芯片、操作系统和网络服务等都还掌握在美国手中，^④ 这种风险就尤应引起决策层的警惕和考虑。

以上风险虽已超出司法制度自身范畴，更多的是基于技术失控所带来的外溢风险，但司法从来不当是点状思维，而是立足社会的系统思维，对此类风险的预估和评判也是司法人工智能极限的另一个方面。

^① 从计算机的发展史来看，其原理也无非将计算理解或等价于确定性的机械运动，只不过这种机械运动从早期的“算珠拨动”发展到了 21 世纪的“电子运动”，而计算的数据量和复杂度也从单一走向多元，类似于有学者所说的从标量（scalar 一个数）到向量（vector 一列数）再到矩阵（matrix 一张二维的数据表）和张量（tensor 三维数据表）的演化。参见王禄生《司法大数据与人工智能技术应用的风险及伦理规制》，《法商研究》2019 年第 2 期。

^② 李本《美国司法实践中的人工智能：问题与挑战》，《中国法律评论》2018 年第 2 期。

^③ 侯猛《互联网技术对司法的影响——以杭州互联网法院为分析样本》，《法律适用》2018 年第 1 期。

^④ 美国信息安全权威专家布鲁斯·施奈尔指出，美国在信息安全技术方面仍然是毫无争议的世界领先者，而且这种领先地位很难被撼动，因为政府的情报预算超过其他国家的同类预算总和。由于互联网起源于美国，其基础架构使得全世界的数据流都会经由美国，即使是两个其他国家之间的数据流动；其他国家的智能产品都在使用美国制造的芯片、操作系统和网络服务，并因此受制于美国法律，“这就是霸权”。转引自 Greg Allen and Taniel Chan, “Artificial Intelligence and National Security”, Belfer Center for Science and International Affairs, Harvard Kennedy School, Harvard University, 2017, p. 97.

五、发展司法人工智能的思路及建议

人工智能司法应用的理论极限，既在于人工智能作为工具的不可能，即不可能超越理性计算范畴、不可能脱离自身条件而发展、不可能超越既定历史条件而创新，也在于司法场景本身的特殊属性，包括司法的不可计算性、动态博弈性和能动创新性，它决定了人工智能在其中的不完全适应性。同时，人工智能对司法的型塑未必是可接受的，伴随着诉讼从线下搬到线上、规则从物理走向网络、司法从解纷扩至治理，对司法程序价值、司法本体价值和技术应用外溢风险的考量和反思正在不断融入“司法规训”当中，以将此种型塑控制在必要范围内或者积极需求替代方案、改进措施等，并减少技术型塑所带来的负面影响。明晰了人工智能司法应用的以上理论极限，就可以避免臆想或妄动，在有所不为的基础上有所作为。

这里不妨据此提出发展司法人工智能的若干建议。在总体思路上，要立足人工智能的工具主义定位，强调司法的主导性、司法的民主性和人工智能在司法场域中的兼容性。具体言之，在对司法性质和属性、价值进行系统反思的基础上，厘定人工智能司法应用的可能和限度，加强司法对人工智能技术的规训。依托司法民主机制，高度尊重人民群众的司法需求、充分发挥人民群众的聪明才智、主动接受人民群众的监督评判，努力实现司法人工智能的可知、可用、可控和可靠。在应用中，要注意司法场景和人工智能的紧密契合，努力做到极限范围内的最大化利用。在具体举措上，司法的主导性、民主性和应用的可兼容性可具体化为以下方面：（1）尝试在最高人民法院或依托国家信息中心设置司法人工智能发展委员会，统一部署和组织实施司法人工智能研发前的司法审查、研发中的监督指导、研发后的评估应用等工作。（2）尝试建立“司法统领、市场运行”的司法人工智能开发机制，由最高人民法院或通过其控制的企业实体（如中国司法大数据研究院）来整体把控司法大数据的安全开发、算法的公开公正，有效避免当事人隐私、意识形态安全等外溢风险，切实解决算法黑箱、算法独裁和算法黑箱等问题。（3）根据具体情况，立足深化司法体制综合配套改革的背景，各级人民法院应依托信息化职能部门与审判部门的衔接互动或法官自主管理委员会等实践机制来真实、高效地汇聚司法人工智能应用方面的具体需求，并对相关的司法人工智能产品提出改进意见和措施等。（4）尝试建立人民法院信息职能部门与司法人工智能开发企业之间的指导监督和交流合作机制，充分发挥法官等司法工作人员“产品经理”“数据标注者”“算法质疑者”等角色功能，以强化司法需求的技术导入。（5）算法的选择和应用均应符合司法人工智能委员会确定的标准。根据现有技术条件，对可以解释的算法应尽可能以公众理解的方式做出充分解释，以击破“算法黑箱”。对于不可解释的算法，应强化事前输入数据质量的控制、不断校正和事后性能的测评、验证，确保其安全有效。（6）加强司法场景的细分梳理，明确人工智能在不同场景中的契合点、可能和限度以及介入的具体方式；在可计算、模式化的司法辅助事务中，应侧重于司法人工智能的替代性介入，辅之以司法工作人员的必要监督；在不可计算的价值判断或涉及能动创新的司法领域中，特别是实体裁判的核心环节，应慎之又慎。要侧重于司法人工智能的辅助性介入和认知性介入，相关的结果输出仅仅作为正面的信息来源或反向的预警参考。

责任编辑：朱志峰