

从数据到知识：智能司法 基础设施的困境反思

陈亮 徐明*

目次

引言	三、要件知识：智能司法基础设施的理想形态
一、数据迷思：智能司法基础设施的建设困境	(一) 要件知识的法理结构——以法律发现为视角
二、法律获取：智能司法基础设施的本质溯源	(二) 要件知识的演绎模式——以法律适用为视角
(一) 为何获取法律——基础设施建设的外在意义	四、知识范式：智能司法基础设施建设的秩序
(二) 如何获取法律——基础设施建设的内在逻辑	结语

摘要 国内现有的智能司法系统普遍采用数据型信息作为基础设施,由于其数据本身存在来源性、模糊性与孤立性的三重困境,致使应用系统难以实现实体性的法律决策辅助。鉴于人机思维的对应性特征,该问题可在法律获取理论的指引下得到解决:智能司法系统的决策思维系发现与适用法律规范的过程,而作为其智识驱动来源的核心基础设施,也应由不确定的法律数据嬗变为确定的法律规则知识。在本体论层面上,原子化的法律构成要件是基础设施的最小知识单元,其互相以数理逻辑符号连接成前束范式的规则表示,且可通过生活经验与司法判决加以解释,以此形成体系化的要件知识网络。而在运行论层面上,该基础设施有赖立法、司法、执法与法学研究主体进行协作建设,其需要构建相互沟通的法律解释共同体,并依据法教义学方法形成法律规则的融贯体系,从而推动智能司法发展迈向知识范式。

关键词 司法人工智能 基础设施建设 法律获取 要件知识 法教义学 计算法学

* 陈亮,上海交通大学凯原法学院博士研究生;徐明,同济大学上海国际知识产权学院副教授、管理学博士。本文初稿曾在第四届“计算法学”国际论坛上进行报告,清华大学徐雨衡博士对此提出了中肯的批评意见;上海交通大学程金华教授与华东理工大学肖梦黎博士为本文提供了宝贵建议,特此致谢。

引 言

现代信息技术为司法场景绘就了不无浪漫的智能化遐想,以数据与算法为驱动的人工智能在法律实践中方兴未艾,并开启了由感知智能到决策智能的跨越式发展。在扑面而来的信息化浪潮下,我国的司法机关也非抗拒技术进步的“新勒德主义者”,自2017年最高人民法院发布《关于加快建设智慧法院的意见》以来,北京“睿法官”、上海“206”与江苏“法务云”等智能化系统层出迭生,其试图为司法活动提供全面的辅助支撑。^{〔1〕}但与之理想目标有所落差的是,此类应用系统的智能化程度尚不如人意,其所能实现的辅助性功能较为有限,并呈现出“重程序而轻实体”的失衡状态:具体而言,大多系统仅聚焦于对流程化事务的辅助工作,主要功能包括线上立案、庭审书记与判决管理等程序性的事项,且功能与功能之间的同质化程度颇为明显;相反,对于案件事实分析、裁判结果预测等具有实质性决策意义的功能,目前只可见于极少数的专用智能系统之中,而其功能的覆盖率与准确率也并不理想。如此种种皆反映了国内智能司法发展的现实瓶颈,人工智能对司法审判的辅助程度与介入空间均显不足,尤其是未能触及核心的实体性法律关系问题,致使其难以在广袤的司法场域下发挥应有的决策支持价值。

究其原因,我国智能司法发展缓慢的症结正是基础设施建设的孱弱。任何智能系统均需特定类型的基础设施作为“燃料”,其一般可通过大数据或其他信息资源的形式加以实现,从而驱动系统整体的运行与输出。然就国内智能司法建设现状而言,相关主体对基础设施的关注程度相对浅薄:其中,司法人员在该问题上基本处于混沌状态,其既未就基础设施本身形成充分的静态认知,也未为基础设施之建设路径提供必要的动态规划;而技术开发人员的工作重心多落于顶层的应用设计之上,其虽知晓底层基础设施的存在,但并不在意基础设施背后的具体机理与逻辑,因而难免产生漠视基本法理或违背司法规律之举。^{〔2〕}据此,《人民法院第五个五年改革纲要(2019—2023)》提出了“深入推进智慧法院基础设施建设”的专项要求,并将其建设目标界定为“全要素”与“一体化”两项具体标准。故着眼于现实需求与未来发展之视阈,理论研究与司法实践均有必要重新审视此类基础设施的建设议题,其不仅具有构建智能司法整体图景的宏观意义,同时亦是实现司法辅助决策的微观驱动与逻辑依归。

有鉴于兹,本文将立足我国智能司法发展的现实方位,对其基础设施的建设问题予以反思。首先,通过系统梳理当前主流智能司法系统的设计思路与运行状况,发现其基础设施所面临的具体困境症结。尔后,从法学理论维度厘清该基础设施的应然本质,将其与上述实然困境进行对照及辨析,指出既有的基础设施形态不适合我国智能司法的发展需要。而在此反思之基础上,本文也将围绕两方面议题做出必要回应:一是“基础设施应当是什么”的本体论问题,即明确我国智能司法的基础设施需转向何种更理想的形态,并进一步剖析其静态的框架结构与动态的演绎模式;二是“基础设施应当如何建设”的运行论问题,即确立相关主体建设基础设施的整体秩序与方法,以期为智能司法的未来发展提供合理的路径参考。

一、数据迷思:智能司法基础设施的建设困境

所谓智能司法的“基础设施”,是指司法人工智能系统运行所依赖的驱动要素,其以信息资源

〔1〕 参见崔亚东:《人工智能与司法现代化》,上海人民出版社2019年版,第88—105页。

〔2〕 参见程金华:《人工、智能与法院大转型》,载《上海交通大学学报(哲学社会科学版)》2019年第6期,第33—48页。

作为最核心的表现形式。^{〔3〕}从世界范围内观之，各国智能司法系统依据信息资源的结构性差异，目前已形成两类典型的基础设施传统：一类是非结构化的数据型信息，如英美国家采用判例文书构造大规模的法律数据库，作为信息要素的文书文本相对独立且相互分离，其不受任何关联性框架的约束；^{〔4〕}另一类则是结构化的知识型信息，如日本学者依据实证法条文构建法律关系知识图谱，其信息要素与要素之间通过符号建立联系，以此表示特定的相关性逻辑。^{〔5〕}正是由于两类基础设施形态的存在，既有的智能司法系统才产生了不同的应用导向：其中，数据型基础设施对应联结主义之智能技术，其擅长从分离的信息要素中自主习得联结性规律，此即形成会预测与创造的虚拟“法律智者”；而知识型基础设施则指向符号主义之人工智能，其借助显式编码与封闭规则的知识表示方法，为司法场景提供了无所不知的自动化“裁判专家”。^{〔6〕}据此可以认为，由信息资源构成的基础设施乃是智能机器的“智识”来源，故在对智能司法系统实施具体开发之前，有必要从源头上为其铺就合理且适法的基础设施，从而方可在应用过程中契合司法规律的理性要求。

当前，我国的智能司法系统主要采用数据型信息作为基础设施。相较于结构化的知识型信息而言，国内司法场景对非结构化的数据积累规模较大，其中最为核心的即是由司法裁判文书形成的文本数据，此外亦包括少量的司法统计数据或派生数据。面对此类基础设施，智能司法系统将依据特定算法而学习其中的司法规律，从而得出由若干要素加权联结的概率模型，并最终应用于特定的功能情景之中。然而，由于文字性的裁判文书不具有天然的可计算性，其他相关数据也多以语言文字作为形式表征，系统在将其转化为数据模型的过程中势必造成信息的失真与损耗。^{〔7〕}也正是受到此番影响，我国司法机关虽已应用了一定数量的智能系统作为庭审辅助，但目前仍未普及所有省市或基层人民法院，且其系统尚存在智能程度不高的“鸡肋”现象。经笔者调研与分析，可将国内智能司法基础设施的既有不足归结为三方面之“数据困境”：

其一，司法数据的来源性困境。可供人工智能学习与训练的资源不仅要求是数据，而且应当是以量著称的“大数据”，由于司法活动所涉领域遍及社会经济生活的各行各业与各方各面，其严格意义上对数据的需求应是接近于无穷大的，几乎所有领域的的数据都有成为智能司法系统驱动的可能性。但是，与该需求形成明显对比的即是当前司法领域内的数据资源现状，无论是信息技术侧的智能系统开发者，还是法律侧的实证法学研究者，各方所掌握的基础资源均依赖于司法活动与司法结果本身的数据公开，其形式基本无法脱离司法裁判文书之范畴。^{〔8〕}虽然近年来我国每年公开的裁判文书数量占结案总量的比例已大幅提升，但其规模仍不能称为“大数据”。事实上，造成此困境的原因并非是司法数据公开程度不高，而是来源过于单一，以至于将其训练成为智能

〔3〕 参见杨建军：《司法数据公开及其程序规制》，载《广东社会科学》2019年第6期，第216—225页。

〔4〕 See Henry Prakken, *A New Use Case for Argumentation Support Tools: Supporting Discussions of Bayesian Analyses of Complex Criminal Cases*, 28 *Artificial Intelligence and Law* 27 (2020).

〔5〕 See Hajime Yoshino, *Logical Structure of Contract Law System: For Constructing a Knowledge Base of the United Nations Convention on Contracts for the International Sale of Goods*, 2 *Journal of Advanced Computational Intelligence and Intelligent Informatics* 2 (1998).

〔6〕 参见宋旭光：《论司法裁判的人工智能化及其限度》，载《比较法研究》2020年第5期，第80—92页。

〔7〕 参见[美] 马克·布尔金：《信息论：本质·多样性·统一》，王恒君、嵇立安、王宏勇译，知识产权出版社2015年版，第104—126页。

〔8〕 参见左卫民、王婵媛：《基于裁判文书网的大数据法律研究：反思与前瞻》，载《华东政法大学学报》2020年第2期，第64—76页。

司法基础设施后,极易产生“因为司法,所以司法”的明希豪森陷阱,即:以司法结果作为智能司法系统决策的学习资料与逻辑模型,并再度依据该模型产生预判的司法结果,以此形成自我复制的无限死循环。由于这一做法总是以既往裁判作为思维基础,其既忽视了法律本身的历史性与进步性,同时也弱化了法律适用的价值逻辑,故最终致使社会影响与利益衡量等个案因素均被排除在裁判之外。

其二,司法数据的模糊性困境。以裁判文书作为智能司法系统的主要基础设施,不仅具有数据来源单一的不足,在数据准确度方面更是存在较大的模糊性短板。裁判文书本身系通过文字形式串联而得的文本,其中用以描述法律与事实的文字在人工智能实现过程中均被归为“名义量”〔9〕,故需通过文本挖掘才可转化为供机器学习理解的数据。较为成熟的做法是对裁判文书进行文本降噪后,再锚定特定信息的具体位置,并采自然语言处理技术对文字内容进行拆分处理,最终依据深度学习模型输出各分词的概率。〔10〕此项计算过程看似符合严谨的科学数理,但在面对灵活的司法实践场景时难免不尽人意。由于法律语言本身就具有明显的模糊性,“同词但不同义”及“不同词但同义”等差异性现象颇为常见,如“占有”或“所有”等用词在刑法语境下与民法语境下会被界定为不同内涵,对此若采同一概率模型则可能与真实情况相差甚远。而即使在单一法律部门的前提下,同一用词也可能显现出迥然意蕴,如民商法理论中的“恶意”一语,既可以表征“故意”之主观状态,也可以用以描述行为人“知情”的客观事实,在司法实践中断不可一语以蔽之。且不论司法活动可能涉及复杂且广泛的社会准则,仅仅在法律语义多元化解释的场域下,若智能司法系统仅单纯以分词的词频作为概率模型的产生基础,而不区分其对应的法律效果,将会大幅增加司法数据的模糊性与不确定性。

其三,司法数据的孤立性困境。与来源性和模糊性等数据内部内容的困境有所不同,既有司法数据的孤立性系其外部形式上的不足。我国智能司法发展正经历着纵深期的精细化推进,〔11〕法律大数据不仅用以完成特定应用之工程任务,亦需要在多个业务部门或领域形成联通效应,以服务其他应用或工程的开发设计。然而现实情况与之相反,我国司法的信息化、智能化进程却陷入了“数据孤岛”的尴尬处境。由于各级人民法院等司法机关均自主开发信息平台或智能系统,各自形成的业务数据均分属于不同的存储载体之中,其形式上的数据格式、通信接口与依赖平台等方面亦无法统一。虽然每一独立的系统内部数据可实现上下游联通与基本一体化,但系统与系统之间的司法数据却面临着壁垒重重、难以流通之障碍。〔12〕故在此视阈下,作为智能司法基础设施的法律大数据在形式上应具有可兼容性之联通特征,如此不只是一是要打通人民法院端与当事人端、律师端等多方面主体之间的数据通道,也需要打通各级人民法院之间的数据上下游链路。这不仅仅是破解数据孤立性的必然的要求,亦可便于从多项端口同时收集智能基础资料,以满足“司法大数据”的数量储备。

然需予指出的是,上述三重司法数据困境皆难以在短期之内得到妥善解决。其中,数据的来源性与孤立性问题受到技术成熟度的制约,司法大数据的规模不可能一蹴而就,数据互通接口也

〔9〕 名义量(nominal data)系与顺序量、区间量及比率量相并列的数据类型,是指无法直接参与数值运算或比较的字符型数值。参见[美]杰弗瑞·希顿:《人工智能算法》,李尔超译,人民邮电出版社2020年版,第30页。

〔10〕 参见王禄生:《论法律大数据“领域理论”的构建》,载《中国法学》2020年第2期,第256—279页。

〔11〕 参见彭辉:《大数据背景下智慧法院发展的回顾与展望》,载高晋康、杨继文编:《迎接智能法学的到来》,法律出版社2019年版,第108页。

〔12〕 参见赵龙、刘艳红:《司法大数据一体化共享的实践探微——以破除“数据孤岛”为切入点的逻辑证成》,载《安徽大学学报(哲学社会科学版)》2019年第6期,第130—138页。

不可能一日即达成,只有经过长期司法实践积累形成互联互通的法律数据库才可解决;而数据模糊性问题更是难上加难,其不仅有赖于语言学研究者建立准确的法律用语语料库,也有待理论与实务研究者逐一梳理法律用语的含义,而后方能形成相对清晰的语义数据模型。在此意义上,克服当前数据困境的唯一方法便是“把一切交给时间”,但伴随着时间的推移与数据的积累,此类基础设施又会落入新的悖论之中:一方面,长期的数据积累可扩大训练样本集中的记录规模,从而使智能司法系统习得更准确的概率模型;另一方面,数据规模的扩大也将使得模型的内容不断改变,在强调一致性的司法场景下反而突显了不确定性,如此难免背离了法律“深度不学习”的本源特征。^[13]于是,国内智能司法基础设施面临着既需优化又无法优化的两难式僵局,在当前形成了一道无法逾越的“数据鸿沟”。

故由此造成的不利后果,便是智能司法系统在实体性辅助决策上的障碍。在国内外法律理论研究者与实务工作者的共识中,智能司法系统的核心要义在于辅助裁判者进行法律决策,^[14]实践中的裁判者也对智能化辅助决策的功能有强烈的需求;对比之下,我国当前主流的智能司法系统仅停留于对机械的程序性工作之辅助,未能深入作为司法内核的实体性法律关系,究其原因正是数据型基础设施的固有不足,从而难以支撑智能系统的整体运行。但即使退一步而论,假设既有的智能司法系统可实现决策辅助之功能,由于上述三重数据困境的存在,其也将在一定程度上有损当事人的程序权利。对于当事人主体而言,其在自动化决策做出前无法了解其中数据的来源与精准度,故不论其主张或抗辩的事实与理由为何者,人工智能始终依据其训练模型给予最终决策结果,^[15]此即是由数据信息的不均衡造成的诉权行使不能。此外,法律实践方法的本质是基于人类理性而对行为的合法性做出评价,^[16]若单一地采用概率数值作为定夺结论,不免有落入“秃头悖论”或“谷堆悖论”之嫌。如此在司法层面上,非感性的“数据鸿沟”会使公民主体的“法律与事实之较量”,不断异化为机械的“技术之较量”。

二、法律获取：智能司法基础设施的本质溯源

面对当前难以逾越的基础设施建设困境,国内的智能司法发展若欲实现突破性进步,则必须从理论层面回溯其基础设施的本源内涵。从根本意义而言,人工智能的运行机制乃是对人类主体的思维效仿,人类主体的裁判思维构成了司法决策的“应然”,而智能机器的辅助决策仅属于“实然”之地位。由该“人机对应”关系可知,智能司法系统所需遵循的准则均取决于人类的思维过程,故对其基础设施的本质判断也需围绕“是否符合人类思维”而展开。对此,首先需要厘清基础设施在智能司法系统中的定位,明确其应当对应人类在决策思维中的何种过程,这便是其基础设施建设的外在意义;尔后,根据人类在该思维过程中的具体要求,确立智能司法系统所需达到的细化标准,以此作为基础设施建设的内在逻辑。在此基础上,通过将这一逻辑路径与国内智能司法基础设施建设现状加以比较,便可推知既有的裁判文书数据型信息是否适格,以及未来的基础设施需作何种必要调整。

^[13] 参见余成峰:《法律的“死亡”:人工智能时代的法律功能危机》,载《华东政法大学学报》2018年第2期,第5—20页。

^[14] See Jack M. Balkin, *The Three Laws of Robotics in the Age of Big Data*, 78 Ohio State Law Journal 1217 (2017).

^[15] 参见马靖云:《智慧司法的难题及其破解》,载《华东政法大学学报》2019年第4期,第110—117页。

^[16] 参见[奥]恩斯特·A. 克莱默:《法律方法论》,周万里译,法律出版社2019年版,第14—17页。

（一）为何获取法律——基础设施建设的外在意义

从技术角度而言,智能司法基础设施的作用在于为司法辅助决策提供依据。与通用人工智能的“资源—逻辑—表现”三层结构相似,智能司法系统一般也包含信息资源子系统、服务支撑子系统与应用子系统三项具体组分,基础设施即对应正是最底部的信息资源层次,其旨在解决“如何驱动系统运行”的原子能力问题。^[17]而系统整体的运行机制正是自底向上的层层抽象,其过程恰好是对人类决策思维的模仿:具言之,人类的决策思维系“根据规范做出选择”,其一般先依主观意愿或客观条件确立特定的规范,尔后再将此规范作为决策行动的准则,最终在若干可能性选项中做出抉择;^[18]智能机器的运行模式也与之类似,其先从底层基础设施中生成特定的逻辑模型,再根据该逻辑模型实施特定的操作行为,进而将操作之结果展示在可视化的界面上。从中不难归纳的是,智能司法基础设施所对应的正是人类决策的规范化思维,其在整体应用系统中的定位应具有准则性意义,即为系统的辅助决策寻找合理的支撑依据。

故从法理角度对应观之,智能司法基础设施的本质乃是实现“法律获取”。德国法哲学巨擘考夫曼曾基于诠释学进路,系统地构建了实质推理模式下的法律获取理论,虽然以类推为中心的等置模式在近年受到些许质疑,^[19]但其中关于“法律发现—法律适用”^[20]的二元视角仍对智能司法发展具有启发式意义。不难理解,此类法律获取过程与法律人工智能的实体性决策过程别无二致:一方面,法律发现技术可用于消解法律疑义、法律漏洞或法律冲突等疑难问题,^[21]并通过法律解释与法律论证方法填补缺省规范,而这一过程的实质即与人工智能的训练或学习过程同理,二者均是从未知或无联系的要素之间建立事实与法律上的可能联系;另一方面,无论是人类裁判者或是机器智能,其均可在梳理所得的完整法律规范链条面前,依据涵摄或等置推理模式完成法律适用项下的任务。基于该人机决策的思维对照可知,由法律发现与适用两环节共同构成的法律获取之作用,就是要求决策主体的法律推理是沿着法律规范的正轨而行进,从而形成早于事实出现的先验性判准。依此对照逻辑可知,智能司法基础设施的目的也应当是获取法律,即获取可供其运行运算的准确规范依据,而该命题可从以下三个维度加以把握:

首先,智能司法基础设施的内容,应是作为法律获取对象的“法律规范”。在法律获取的理论中,法律规范是以“应然性”作为基本特征的,其与“实然性”的案件事实存在着复杂的张力关系:从异质性的角度而言,规范与事实之间存在着无法跨越的隔阂,法律获取的目光仅能在二者外部进行往复流转,但永远无法突破障碍而在内部加以调和,此即是“休谟问题”所揭示的因果鸿沟;^[22]而从同化性的角度言之,应然性的法律规范也具有向实然事实开放的可能性,法律规范依据案件被划分为“构成要件”,案件本身亦依据法律规范形成“案件事实”,二者在涵摄发生前处于相互塑

[17] 参见[美]戴维·克伦克、兰德尔·博伊尔:《管理信息系统》,冯玉强等译,中国人民大学出版社2019年版,第240—242页。

[18] 参见宾凯:《从决策的观点看司法裁判活动》,载《清华法学》2011年第6期,第94—108页。

[19] 此类观点认为考夫曼所持立场为不彻底的诠释学,致使类推理论在形式推理与实质推理的断裂局面之间徘徊,无法确认其理性或非理性之状态。参见陈辉:《从诠释学立场对考夫曼推理体系的分析和完善》,载《法制与社会发展》2014年第1期,第136—146页。

[20] 参见[德]阿图尔·考夫曼:《法律获取的程序:一种理性分析》,雷磊译,中国政法大学出版社2015年版,第27页。

[21] 参见王洪:《制定法推理与判例法推理》,中国政法大学出版社2016年版,第162页。

[22] 参见[美]撒穆尔·伊诺克·斯通普夫、詹姆斯·菲泽:《西方哲学史》,邓晓芒、匡宏等译,北京联合出版公司2019年版,第279—280页。

造之状态。^[23] 据此不妨认为，法律规范是在时间之流中的动态对象，其在决策进行前可由特定事实等置而来，在决策进行时其便是应然性的大前提，而在决策后又将成为未来决策的可能依据。在此意义上，由于基础设施在智能司法系统运行前即需建设完毕，目的在于为系统提供在“决策时”可遵循的准则，故无论是基于符号主义抑或是联结主义的技术立场，其系统底层的基础设施均应恒具应然性。

其次，智能司法基础设施的来源，应是作为法律获取场所的“法律渊源”。法律获取的过程不仅需要获取法律规范，更需要在特定的场所中开展获取活动，这就要求智能司法基础设施不得超出一国法律渊源的范畴。由于法律渊源一般是指“有效力的法律表现形式”，其强调法对特定社会事实的拘束力，^[24]故对应的基础设施内容应主要围绕立法规范而展开；当然，考虑到司法解释或指导案例等非立法规范也可为司法裁判提供依据，此类法律渊源的概念范畴也可作适当扩张，然则无论采用何种理解，智能司法基础设施均需具备最低程度的裁判效力。不仅如此，法律渊源亦是决定该基础设施形态与结构的关键要素：具体而言，制定法规范具有相对清晰的逻辑，不仅在规范内部各要件之间具有微观联系，在规范与规范之间也有宏观的体系化联系，其即对应为知识型基础设施；相比之下，判例法规范的独立性相对较强，尽管各判例内部具有一定程度的逻辑牵连，但判例与判例之间的体系性明显不足，因而其基础设施应采用离散的数据型信息加以表示。由此可见，在建设智能司法基础设施的过程中，开发者需先行寻找在该国具有裁判效力的渊源依据，尔后进一步确定信息资料的具体呈现形态，如此方能作为人工智能系统的合理驱动。

最后，智能司法基础设施的功能，应是通过法律获取而补强决策过程的“可解释性”。如前所述，基础设施需为智能司法系统生成特定的逻辑模型，应用系统正是依据此类模型才能做出相应的决策。但犹如国内外学者所指出的那般，任何人工智能应用的决策模型均是一类不确定的“黑盒”，其内部结构或运行结果之归因皆无法被外界解释，甚至难以为系统开发者所解释，致使智能司法辅助决策仅有结论而无过程，从而极易诱发不合理、不公正的判决出现。^[25] 对此，笔者反倒认为不必过于悲观，决策过程的不可解释性乃是司法场域下的常规现象，即使是由人类主体所担任的司法裁判者，其目光在事实与规范之间往返流转的机制也无法被完全清晰地认知或描述，这便是实践中对相似案情得出差异化裁判的根本缘由。而如此裁判之所以能被社会所接受，原因正是在于法律获取理论之支持，即：不论司法决策的过程呈现出何种复杂性，只须裁判者的决策依据系准确的法律规范，其所得之决策结论就是符合司法理性的产物。因此对智能司法系统而言，开发者并不需要过多关注决策模型本身的构造，只要采用法律规范作为系统的“心智”起点，其便可有效控制决策模型在暗默空间下的恣意生长。故正似信息论奠基人香农所语，信息具有“消除随机不确定性”^[26]的作用，由信息资料构成的智能司法基础设施也应具备确定性之特征，其在司法辅助决策中的角色担当至少有二：一是解释真实且准确的法律规范，作为后续运算或决策的逻辑准绳，该过程即与法律获取的“发现”视角相对应；二是解释规范与事实之间的涵摄或等置机制，使

^[23] 参见[德]阿图尔·考夫曼：《法律获取的程序：一种理性分析》，雷磊译，中国政法大学出版社2015年版，第157—158页。

^[24] 参见刘作翔：《“法源”的误用——关于法律渊源的理性思考》，载《法律科学（西北政法大學學報）》2019年第3期，第3—12页。

^[25] See Reuben Binns, *Algorithmic Accountability and Public Reason*, 31 *Philosophy and Technology* 543 (2017).

^[26] See Claude E. Shannon, *The Mathematical Theory of Communication*, 27 *Bell System Technical Journal* 379 (1948).

外界可以观测法律推理的展开情况,此则与法律获取中的“适用”过程形成对应。

(二) 如何获取法律——基础设施建设的内在逻辑

毋庸讳言,我国当前智能司法基础设施的建设逻辑,主要是参考了英美法系国家的法律获取进路。就世界范围内的智能发展进程而言,较为发达的智能司法系统主要集中在以判例为法律渊源的英美国家,^[27]其大多借助类比先例之推理模式对未知的案件事实做出判断或评价。正如美国大法官霍姆斯有关“普通法生命”的经典论述,^[28]判例在英美国家是阐释立法规则与填补法律漏洞的实证手段,其不仅包含司法适用的实体规则与程序规则,亦同时涵盖了规范逻辑与价值判断。事实上,此类以联结主义为主导的智能司法系统也存在人机思维的对应关系:其中,英美法官的法律获取过程是将待判事实与既往案例在形式上加以比较,从而检索得相同或最为相似之先例,并受其判决结果之参照约束;受此影响,智能司法系统普遍采用判决文本数据作为基础设施,通过生成的经验性概率模型,使未决的案件事实受其约束。^[29] 不难理解,此类基础设施建设逻辑虽是实现司法辅助决策的合理行径,但由于驱动系统应用的“智识”来源系海量的数据型法律信息,故其适用条件仅限于具有大量判例积累的英美法系国家。相比之下,我国借鉴该模式而实现的智能司法系统并不理想,且其存在着两方面的逻辑错位:

一方面,我国的裁判文书数据未触及法律规范本身,难以解决法律发现之问题。就法律发现的目标而言,无论是遵循法律形式主义或现实主义的立场,司法过程均需将探明具体的法律规范作为后续适用之基点,其旨在达到抑制非理性裁判的目的。然颇为遗憾的是,我国智能司法系统尚未实现这一目标,尽管其与英美国家相关系统均采用裁判文书作为基础设施,但此类文书数据在不同场域下的规范性意义并不相同:其中,英美法系国家的裁判文书系作为法律渊源的“判例”,其文本数据是对既有事实的“实然评价”,故可在后验案件中得到直接适用;而在国内司法场景下,裁判文书仅是“案例”的作用,其难以实现对规范本身的“应然解释”,其法律的适用效力主要来源于制定法规范。换言之,我国智能司法系统所依赖的裁判文书并非是一类“法律规范”,其仅是一项经验性的“社会事实”,从而难以逾越休谟问题所揭示的因果鸿沟。即使是可作为裁判依据的最高人民法院指导案例,其在我国司法体系下的适用地位也仅及于“参照”而非“约束”,^[30]况且审判人员适用指导案例的本质乃是参考案例背后的法理逻辑,亦非如联结主义智能一般对裁判文本的相似度进行比较。在此意义下,由裁判文书数据驱动的法律发现过程未效仿人类的“找法”逻辑,此类将特定的实然结果映射至另一结果上的机械推演,终究不足以跳脱法律适用中的语义模糊性窘境。

另一方面,国内裁判文书积累相对匮乏且存在分歧,也难以实现法律适用之目的。英美法系国家的司法判决均受先验判例之约束,因而类案事实均可得相同或相近之裁决。然而,我国司法场域下的案例是对特定事实的实然诠释或评价,其中势必包含着裁判者对不同规范的主观认知与选择,更何况对同一规范亦可能呈现多元化、非确定的解读,故司法实践中“同案不同判”的现象尤

[27] 早在1991年加拿大就开发了以司法案例文本数据库为基础的专家辅助裁判系统,而后至今美国的COMPAS、HYPO等法律风险衡量与评估应用亦沿用了这一算法。参见[美]凯文·D.阿什利:《人工智能与法律解析:数字时代法律实践的新工具》,邱昭继译,商务印书馆2020年版,第101—104页。

[28] 参见[美]小奥利弗·温德尔·霍姆斯:《普通法》,冉昊、姚中秋译,中国政法大学出版社2006年版,第1页。

[29] See Marcello D. Bello & Bart Verheij, *Evidence and Decision Making in the Law: Theoretical, Computational and Empirical Approaches*, 28 *Artificial Intelligence and Law* 1 (2020).

[30] 参见孙跃:《指导性案例跨类型适用的限度与进路》,载《交大法学》2020年第1期,第166—178页。

为明显。^[31] 而此番情形反映在智能司法辅助决策上的痛点即是，若将纯粹的裁判文书数据直接列为基础设施，则法律获取的对象本身就存在着严重的自我冲突，这不仅在司法价值上有违裁判的确定性要求，在实现效果上亦会大幅降低决策的可信效度。但对于这一决策信度的不足，国内部分观点仍以普通法的法律获取路径为指南，^[32]其尚未意识到问题的本质与核心，而将关键症结归于裁判文书数量或质量的短板，并试图通过强化数据库记录的方式提升决策的精准度。殊不知统计学上的“安斯库姆四重奏”^[33]模型已揭示了其可能的反例，如此不但未能达到缩减误差的预期效果，反而加剧了前述三重困境的迭生。

依此逻辑，我国智能司法基础设施的内容应转向制定法规范，其基础设施的形式也需转向知识型法律信息。与裁判文书数据的联结主义立场相对应，制定法规范因关联了不同的构成要件元素，故其形式上属于符号主义学派语境下的“知识结构”。^[34] 在功能定位上，该结构以知识的表示与推理为核心，其中“表示”功能在司法场景下即对应法律发现之过程，而“推理”则指向法律适用的程序。因此，相比裁判文书数据在前述两方面的不足，此类制定法规范知识与法律获取过程的基本要求高度契合：一方面，我国将制定法规范作为主要的正式法律渊源，且其内涵与外延可通过法律解释的方法得以厘定，从而为智能系统提供“找法”依据；另一方面，制定法规范具有权威性与稳定性，以其为核心的实证法律信息可消除法律实践中的机会主义现象，^[35]而这正是作为智能系统的“用法”逻辑。不仅如此，在我国智能司法基础设施建设的目标层面上，依据制定法规范及其解释构造的知识型法律信息，也有超越裁判文书数据型信息的现实意义。个中理由如下：

其一，基于制定法的法律规范知识可覆盖法律“全要素”领域。以智能司法系统实现辅助决策的基本意义，是使法律评价自动地投射至纷繁复杂的社会事实之上，其必然要求作为“心智”来源的基础设施应尽可能地覆盖全面。倘若抽象地将社会生活事实视作一个平面，成文的制定法规范依据其知识结构，即可在平面上形成稠密的“法律关系网”。这一网络由不同的法律构成要件连接而成，网络上的每一结点对应一类确定的法律关系模型，且结点模型之间依据规范解释的方法而联系紧密，特定的社会事实若不符合其中的某一模型，也必定可对应至另一确定的模型。相比之下，裁判文书数据型法律信息在此平面上的表征仅是若干离散的点阵，各份文书对应的司法案例之间的分界相对清晰，较少存在法律关系交叉或过渡地带。事实上，真实的社会场景并非是如数据型信息般“非此即彼”的离散模式，其涉及的法律问题往往是“此中有彼”的复杂形态，这便要求基础设施提供一体化的网络式支撑信息。故就未来智能司法系统的辅助决策需求而言，其基础设施更宜回归法律获取对象的规范本身，依据客观制定法及其主观解释，编织社会法律关系的全景地图。

[31] 参见阮堂辉、陈俊宇：《“同案不同判”现象的内在逻辑与治理路径》，载《学习与实践》2018年第7期，第71—77页。

[32] 参见涂钊：《美国司法数据应用的过去、现实及争议——兼论对我国司法大数据应用的启示》，载《科技与法律》2020年第1期，第57—67页。

[33] 安斯库姆四重奏(Anscombe's quartet)是指由四组统计特性基本一致的数据，绘制出具有显著差异性的图像，其体现了用数据与算法模拟真实事件时的机制缺陷，用以证明离群值对数理统计的影响之大。参见[美]杰弗瑞·希顿：《人工智能算法》，李尔超译，人民邮电出版社2020年版，第149—151页。

[34] See Michael A. Gilbert, *Informal Logic, Argumentation Theory and Artificial Intelligence*, 22 *Informal Logic* 12 (2002).

[35] 参见[德]古斯塔夫·拉德布鲁赫：《法哲学入门》，雷磊译，商务印书馆2019年版，第35页。

其二,基于制定法的法律规范知识可达到功能“一体化”标准。“一体化”是要求智能司法系统应覆盖“发现—适用”的法律获取全过程,且可能同时辐射至法律适用决策之后的影响与社会效应。^[36]就此意义而言,智能司法基础设施不仅需要深入剖析规范本身的内容要件与形式结构,亦有必要根据法律价值的导向指引而自我更新,最终呈现为法律规范之完善或解释论上的优化。尽管裁判文书数据型信息也对法律价值的判断有所涉及,^[37]但其效果也仅限于“个案正义”,而无法在相异的司法场景下普遍适用。况且数据型法律信息知识的迭代难度较高,当法律规范因修订或废止而更新时,原有的文书数据将失去法律获取的参照意义,致使其相关的基础设施只得重新积累。对此,以制定法为中心的知识信息只需调整内部要件的组成结构,即可直接实现法律规范之跟进。更为重要的是,法律智能系统的“一体化”目标也不局限于司法场景,基础设施之效用也应贯穿立法、行政与监察等社会治理的立体过程,^[38]而其均将制定法规范作为“依法”治理的基石。故从法律活动的融通性角度加以考量,此类知识型基础设施具有串联法律制定及实施等工作的启迪意义,其未来实践或可突破当前司法数据之孤立性困境。

当然,倡导建设以制定法知识为核心的基础设施,并不意味着裁判文书数据被我国智能司法系统完全舍弃,而是应明确此二者的主次逻辑。虽然如前所述,单纯的数据型信息无法完成法律获取的艰巨任务,唯有制定法及其解释才是法律规范的形式构成;但不可否认的是,经验式的案例裁判文书也涵盖了法律的目的理性与开放逻辑,^[39]其可辅助特定事实上的价值推导。易言之,制定法规范与既有案例均在人类裁判者的司法思维中有所侧重,智能司法基础设施的建设即可结合两类法律信息的优势,使之在不同的技术层次上展开。当前国内外部分司法人工智能系统已有类似实践,^[40]其将成文的制定法作为应用系统的主要法律获取对象,旨在为人工智能解决“依据何法律”的基础问题;至于法律发现与适用后的效果命题,应用系统则试图从裁判文书数据信息中获取校验规则,从而作为智能司法辅助决策的价值判断。具体而言,由于既往案例对相似的特定事实可能给予不同的法律评价,且此类结果的差异性程度一般较弱,故可认为类案判决提供了一段“价值区间”,若依据制定法推理所得的结论落入该合理范畴之内,其即为契合社会价值导向的司法裁判。在此意义下,规范的知识适用与裁判文书的数据校验存在主次之别,驱动智能司法系统的核心基础仍需深入法律规范条文本身,通过梳理全面的法律知识信息联通网络,以达到使机器智能获取法律的目的。

三、要件知识：智能司法基础设施的理想形态

在本体论之层面上,我国的智能司法基础设施需从数据型信息嬗变为知识型信息。尽管国内

[36] 参见胡铭、张传玺:《人工智能裁判与审判中心主义的冲突及其消解》,载《东南学术》2020年第1期,第213—221页。

[37] 参见朱晖、刘晨晖:《大数据在同案审判中的应用研究》,载《法律适用》2019年第20期,第47—54页。

[38] 参见雷磊:《中国特色社会主义智慧法治建设论纲》,载《中共中央党校(国家行政学院)学报》2020年第1期,第99—110页。

[39] 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》,载《法商研究》2019年第2期,第101—112页。

[40] See Rohan Nanda, Giovanni Siragusa, Luigi Di Caro, Guido Boella, Lorenzo Grossio, Marco Gerbaudo & Francesco Costamagna, *Unsupervised and Supervised Text Similarity Systems for Automated Identification of National Implementing Measures of European Directives*, 27 *Artificial Intelligence and Law* 199 (2019).

已有部分学者认识到此项逻辑转型的重要性，并建议以法律关系构成要件构建司法人工智能的知识网络，^{〔41〕}但在实操层面上仍需进一步明确具体的要件知识抽取与组方法。具体而言，法律获取可拆解为法律发现与法律适用两项既相对分离又互相关联的过程，其对于要件知识网络的搭建亦具有启发意义：一方面，借助法律发现理论可剖析法律规范之具体结构，解决基础设施的知识形态“应当如何构成”的静态问题；另一方面，法律适用则可实现从法律规范向案件事实的映射，解决此类知识“应当如何演绎”的动态问题。

（一）要件知识的法理结构——以法律发现为视角

从人机思维对应的角度而言，我国的智能司法基础设施需以“要件知识”作为基本结构。要件知识是在“要件事实型民事裁判论”^{〔42〕}视角下，根据各法律关系的不同构成要件，相互关联组合成为特定法律事实的知识网络。与人类司法裁判者所采用的“要件审判”^{〔43〕}思维一致，人工智能在判断案件事实是否构成特定法律关系时，亦需逐一判断法律关系的各构成要件成立与否。因此，梳理要件知识不仅需要遍历全域法律关系的各构成要件，亦需要从构建体系化的法律规范“知识库”的角度出发，明确要件之间的排列组合及其与不同法律关系的对应模型，进而满足本体表示、信息生成、交互输出等多方面技术的集成化要求。^{〔44〕}这一基于要件“本体”与要件之间“联系”的路径，即与以笛卡尔为代表的理性主义哲学家所提倡之“知识组方法论”^{〔45〕}同源，而当前国内外智能应用设计所用之知识表示方法亦是以该描述逻辑为主流。^{〔46〕}在此借鉴意义下，对作为司法人工智能基础的法律规范信息进行解构时，也需要从内部结构与外部联系加以把握。

首先，“法律要件”是法律规范作为知识型信息的内部最小单元。拉伦茨将法律规范的核心聚焦于“构成要件”及其对应法律关系的“法效果”，立法者通过权利与义务的设置而引导主体可为、应为或不为某类行为，^{〔47〕}该逻辑下的法效果之归因即是特定构成要件的组合与事实相该当。因此，智能司法系统若欲从法效果角度做出合法或非法之决策选择，其判断依据正是案件事实是否同时符合特定法律关系的所有构成要件，而通常做法便是罗列出可能与纠纷案件相关的一切法律关系模型，并逐一拆解至最小单位的构成要件与事实进行比对。不难理解，该过程中的关键问题在于如何尽可能细致地获取法律构成要件，对此建议参照数理逻辑上之“原子命题”建立法律规范的“原子要件”列表，即：将特定法律规范条文等来源中的表述进行简化分解，直至作为最终呈现结果的各要件不能再被分解为更简单的要件，甚至不包含各类表示否定、并列、选择等连接词，从而

〔41〕 此类观点基于人工智能的符号主义学派立场，证成了以形式逻辑与数理逻辑为驱动的机械唯物智能参与司法活动的可行性。参见高翔：《人工智能民事司法应用的法律知识图谱构建——以要件事实型民事裁判论为基础》，载《法制与社会发展》2018年第6期，第66—80页。

〔42〕 要件事实型民事裁判论是指以实体法律规范的要件结构为基础，将案件事实对应地拆解为相关的构成要件，以此作为民事诉讼审理依据的民事裁判方法。参见许可：《民事审判方法：要件事实引论》，法律出版社2009年版，第2页。

〔43〕 参见邹碧华：《要件审判九步法》，法律出版社2010年版，第92—108页。

〔44〕 参见周佑勇：《智能技术驱动下的诉讼服务问题及其应对之策》，载《东方法学》2019年第5期，第14—19页。

〔45〕 参见〔日〕野中郁次郎、绀野登：《创造知识的方法论》，马奈译，人民邮电出版社2019年版，第45页。

〔46〕 See Georgios Kourtis, Evangelia Kavakli & Rizos Sakellariou, *A Rule-Based Approach Founded on Description Logics for Industry 4.0 Smart Factories*, 99 IEEE Transactions on Industrial Informatics 1 (2019).

〔47〕 参见〔德〕卡尔·拉伦茨：《法学方法论》，黄家镇译，商务印书馆2020年版，第319—322页。

逐一编号便可形成完整的要件列表。如表 1 所示,就是依据我国《商标法》第 13 条关于复制、摹仿、翻译他人驰名商标的法律规范所建之原子法律要件列表。需要指出的是,原子要件的“发现”来源并非局限于制定法条文,包括司法解释、法规政策等,甚至特定情形下的学理解释亦是此类法律要件的重要来源。刑法中的“犯罪构成”问题即是一则典例,我国刑法典及其司法解释未明文规定“犯罪”概念的构成要件,而在司法实践中无论是采“要件论”抑或“阶层论”,均不可否认其是依据学理解释之通说进行的判断。

表 1 我国驰名商标保护规范涉及的原子法律要件

要件	含义	要件	含义
a	申请人商品与他人商品相同	g	他人在中国注册驰名商标
b	申请人商品与他人商品类似	h	所申请商标与他人驰名商标混淆
c	申请人申请注册商标	i	所申请商标误导社会公众
d	所申请商标是复制他人驰名商标	j	所申请商标致使他人权益受损
e	所申请商标是摹仿他人驰名商标	k	所申请商标准予注册
f	所申请商标是翻译他人驰名商标	l	所申请商标可以使用

其次,法律要件之间依据“法律规则”形成相互联系的外部知识网络。在阐述此项命题前,不妨先以上述之驰名商标的保护规范为实例,从具象的微观法律发现层面予以考察。我国《商标法》第 13 条第 2 款与第 3 款分别系在同类商品或不同类商品上申请注册与驰名商标近似的商标之禁止规范,借以上述原子法律要件列表,可将严格的智能知识逻辑表达如下。其中,将全称量词“ \forall ”或存在量词“ \exists ”等修饰置于表达式前部,用以表达“一定”或“可能”等语义;式中以单条件的蕴含逻辑符“ \rightarrow ”联结规范中的“行为模式”与“法效果”,而此两部分内部则通过否定符“ \neg ”、析取符“ \vee ”或合取符“ \wedge ”进行串联,以表征各原子要件之间的逻辑联系:

$$\forall c \exists h [(a \vee b) \wedge c \wedge (d \vee e \vee f) \wedge \neg g \wedge h \rightarrow \neg k \wedge \neg l]$$

$$\forall c \exists j [(\neg a \vee \neg b) \wedge c \wedge (d \vee e \vee f) \wedge g \wedge i \wedge j \rightarrow \neg k \wedge \neg l]$$

巧合的是,这一可供智能司法系统直接理解的知识网络,与法律规则的“假定—处理—制裁”三要素结构具有高度相似性。假定要素为法律规则的适用框定了必要的条件,其与一阶逻辑上之量词表示具有同理;而处理要素与制裁要素实质上分别对应法律规范中的行为要求与效果,可类比至逻辑式蕴含符号两侧的组成部分。之所以选择上述一阶谓词逻辑或霍恩逻辑的表达式作为要件知识的网络组成,原因乃是基于其结构严密性与稳定性之考量,以此体现智能司法系统“深度不学习”的本质要求。尽管该基础设施可能存在无法表示不确定性知识的缺陷,^[48]但其并不影响法律获取过程之实现,可供智能司法系统学习的知识型信息本身即是稳定的法律规则,而这也恰

[48] See Cozman F. Gagliardi & Maua D. Deratani, *The Finite Model Theory of Bayesian Network Specifications: Descriptive Complexity and Zero/One Laws*, 110 International Journal of Approximate Reasoning 107 (2019).

恰印证了法律规范的准确性追求。因此，借鉴一阶逻辑中的“前束范式”^[49]，实然法背景下的法律要件知识网络可统一地表述为以下逻辑式之形式，原子要件也就据此建立了相互之间的联系，其对应的法律规则之意义即为“针对当然的或可能的行为 x ，若符合法律关系 $F(x)$ 的所有构成要件，则得到 $G(x)$ 的法律效果”。

$$\Box x[F(x) \rightarrow G(x)]$$

最后，法律要件的内涵与外延需通过非要件知识的其他信息予以确定。无论是作为整体的法律规则，还是作为最小单元的法律要件，其在法律规范等效力来源中的文字表述往往是模糊而抽象的，如表 1 中要件 i 有关“误导”的判定就难以通过文字本身进行简单描述。故对于人类主体或机器智能而言，其司法决策活动必然需要结合一定的社会现实加以理解，这就需要明确界定各要件的内在意义与外部范畴。据此建议通过两种方法予以完善：一是法律解释的方法，即对各原子要件的意义以及各要件之间的逻辑进行解释，使得卷帙浩繁的法律规范在面对多元化的社会纠纷时，可通过语义、体系与目的等多类解释方法回答“要件可能映射至何类社会事实”的知识本体问题，以及“要件之间应通过何类连接词组合”的知识关联问题，并逐一选择其中之“最佳解释”^[50]；二是经验参照的方法，即将既往的司法案例视为法律规范在社会生活中的秩序指引，在归纳不同经验事实之法律共性的同时，通过梳理复杂而多元的判例内容与细节，用以填充各法律规范要件可能的映射范围。需予指出的是，此两类方法正是同一问题的双重面向，司法审判经验本身就是法律解释的另一类表现形式，裁判文书数据对于人工智能而言，其作用并非构造作为判准的法律规则，而是基于实然经验对规则本体进行解释。因此，智能司法基础设施的核心尚未脱离规范之范畴，其仍是以要件知识为表现形式的法律规则。

（二）要件知识的演绎模式——以法律适用为视角

关于法律适用的演绎推理与人工智能思维过程的同源性问题的讨论。^[51]而在以要件知识网络作为智能司法基础设施的背景下，不难发现前述之一阶逻辑的前束范式与演绎推理的“三段论”模式高度相似：大前提正是前束范式规则本身，当待判案件事实符合其小前提“ $\Box xF(x)$ ”时，便可经涵摄推出 $G(x)$ 之结论。事实上，这一涵摄过程更宜对应地理解为智能司法辅助决策的“分类”算法，^[52]即利用给定预期法效果的输出规则，对未知社会生活事实的法效果加以预测的“监督学习”方法。^[53]基于此视角，智能司法系统的法律适用模式也就自然地同时包含了应用与积累基础设施的两项功用，以下具言之：

其一，智能司法系统的法律适用是对知识型基础设施的应用过程，其实质是实践法律规则的逻辑。出于对形式正义的价值追求，法律适用就应是由事实认定与实质推理所构成的结合

[49] 前束范式(prenex normal form)是使用谓词逻辑所描述形式语言的标准范式，其表达式中之量词均位于式首，以约束延伸到表达式末端的全部作用域。参见左孝凌、李为鑑、刘永才编著：《离散数学》，上海科学技术文献出版社 1982 年版，第 73—75 页。

[50] 参见孔祥俊：《法律解释与适用方法》，中国法制出版社 2017 年版，第 207—210 页。

[51] See Bruce G. Buchanan & Thomas E. Headrick, *Some Speculation about Artificial Intelligence and Legal Reasoning*, 23 *Stanford Law Review* 40 (1970).

[52] 参见陈亮、郭佳雯、武建功、王占全、史令：《基于法计算学理论的人工智能辅助决策算法研究》，载《华东师范大学学报(自然科学版)》2019 年第 5 期，第 85—99 页。

[53] See Neda Sakhaee & Mark C. Wilson, *Information Extraction Framework to Build Legislation Network*, 29 *Artificial Intelligence and Law* 35 (2020).

体。^[54] 当以符号主义为主要立场的人工智能进行辅助性的法律适用时,其所关注的是如何依据确定的知识要件网络规则推理得出确定的法效果结论,即智能司法基础设施在社会实践中的延展性问题,而非基础设施的内部构造问题之本身。抽象地概括该适用过程便是,事先根据法律规则确定 $F(x)$ 构成要件的组合,且每一组合均对应一项法律效果的“分类模型” $G(x)$,最终将输入的事实与 $F(x)$ 组合的列表进行比对,寻求相等或最为近似的一项组合记录,以输出其对应的 $G(x)$ 模型。在此过程中,法律适用需要基于要件知识的逻辑结构展开三轮涵摄:其中,第一轮是量词之涵摄,解决法律规范中“可能”“应当”或“可以”等多元量词与单一案件事实之间的对应问题,证成法律规则的假定部分 $\square x$;第二轮是构成要件之涵摄,解决法律逻辑规则中以合取符并联的各构成要件与案件事实各要素之间的对应问题,证成法律规则的处理部分 $F(x)$;第三轮是法律规范之涵摄,即最终依据前两轮涵摄之结果,决定是否推理得出预设的制裁部分 $G(x)$,从而环环相扣地得出辅助决策之结论。由此可见,若将要件知识作为智能司法基础设施,应用系统的法律适用行动便能充分体现法律推理的确定性;同时,由于法律适用仍是以法律发现的知识结构为展开,此类一以贯之的逻辑模式也从侧面证实,基础设施由数据型到知识型的转型具有显然的可行性与实践价值。

其二,智能司法系统的法律适用是对数据型基础设施的积累过程,其实质是解释法律规则的要件。人工智能系统在进行每一次法律适用的运算时,均会对输入的案件事实信息加以记录,并作为一条新的数据条目加入既有的基础设施之中。尽管前文已提及,数据型信息不宜作为我国智能司法系统的主要智识来源,然其不意味着数据条目的积累在基础设施建设中不具有任何价值。相反,该举措之意义有二:一者,不同社会生活情境下的法律与事实数据,本质上就是通过既往案例解释法律规则的各项构成要件,虽然此积累过程不触及要件知识本身,但仍可通过“经验”的方式深化对知识本体与联系的理解;二者,利用积累所得的数据型信息,可实现对司法裁判“全要素”的分析与统计,如此既能便于人民法院等主体直观地了解案件的具体情况,也有助于从结果角度回溯审查该决策结论的合理性。因此,智能司法系统有必要按“ $\square x F(x)$ ”的形式分别记录各法律构成要件该当与否,其不仅是供人工智能再度学习与优化的经验参考,更是抑制法律适用不当情形的校验规则。

由此可见,智能司法基础设施具有动态性之特征。尽管由要件知识规则构成的基础设施具有相对稳定的网络结构,但稳定性并不等同于绝对的静止性,尤其是就法律适用的意义而言,此类基础设施至少具有两方面的动态表征:一方面,要件知识网络需因制定法的改变而变化。无论是作为法律效力渊源的立法规范,抑或是指导性案例、司法解释等裁判依据,其本身必然伴随着规则的新设、修改与废止,而该规则内容的变化也将致使 $\square x$ 、 $F(x)$ 或 $G(x)$ 部分的结构变动,为此需要在智能司法基础设施中生成新的规则。另一方面,要件知识网络也需要根据社会情势的变迁而做出必要调整。正如法律获取理论所揭示的那般,应然性的法律规范需向实然的案件事实开放,^[55] 当上述数据型的基础设施积累至一定规模或契合特定程序要求时,其也需要被人为地固定为结构化的要件规则知识。在此意义上,对智能司法基础设施的建设既需要协调体系内部的法律规则结构与冲突,同时也需要协调体系在外部实践运行的社会效应,这便是通过基础设施自身的不断迭代与升级,促使智能司法辅助决策真正深入公平与正义的内核。

[54] 参见[英] 尼尔·麦考密克:《法律推理与法律理论》,姜峰译,法律出版社2018年版,第103页。

[55] 参见[德] 阿图尔·考夫曼:《法律获取的程序:一种理性分析》,雷磊译,中国政法大学出版社2015年版,第185—187页。

四、知识范式：智能司法基础设施的建设秩序

伴随着智能司法基础设施之本体转向知识型信息,其基础设施的建设秩序也需升级为对应的“知识范式”。在既往之“数据范式”下,以裁判文书文本为主的数据型基础设施面临着三重建设困境,且摆脱困境的唯一方法仅是通过时间推移而不断积累数据,此即致使智能司法系统在短期之内难以发展。与之有所不同的是,知识型基础设施的建设要义并非是寻求浩如烟海的司法素材,其重心在于构建一套成体系化的“要件知识”标准,这既是基于法律发现的规则构成之标准,也是基于法律适用的事实解构之标准。易言之,未来的智能司法基础设施建设不应有赖于时间的被动演化,其应当转向主体能动性之范畴,即由人类主体主动构造作为基础设施的要件知识。但需予指出的是,由于法律规范所涉范围极度广泛,单一主体难以完成如此庞大的构造任务,其不得不由多元化的主体进行分工构建;而倘若由不同主体进行共同设施,考虑到各方对法律规范可能存在不同理解,由其生成的基础设施也必然存在分歧。故在此番知识范式下,基础设施建设所面临的关键问题便是主体层面之“秩序”,其不仅需协调不同主体之间的认知差异,也需明确各主体之间的相互协作关系,从而使所得的基础设施可贯穿智能司法系统的全域应用场景。

事实上,智能司法基础设施在知识范式下的建设秩序问题,可从法教义学的方法论层面寻得线索。不难理解,法教义学方法与知识型基础设施具有天然亲缘性,前者系法律形式主义立场下的产物,而后者乃是符号主义人工智能的驱动基础,二者均以确定性推理与紧密性逻辑作为核心要素。就此逻辑观之,法教义学的知识生产机制或可为智能司法基础设施的建设秩序提供启示,即:各方应通过法律条文或相关的法理梳理教义化的法律规则,并再度依据规则教义而实现其知识本体的生成与发展。^[56] 不仅如此,依法教义学进路构建基础设施更有两方面的可行性基础:一是从国内现实考虑,法教义学乃是学术研究 with 司法适用的主流方法,特别是在当前国内部门法体系愈臻强化的前提下,法教义学在司法场景下相当普遍的指导价值;^[57]二是从域外经验的角度言之,以制定法为法律渊源的国家皆强调法教义学对司法智能化的意义,如日本智能司法系统的开发始终遵循法律规范分析的进路,从而在社会层面形成了规模化的学术与技术团体。^[58] 对此,笔者根据法教义学方法论的运行要求,为我国智能司法基础设施的建设问题提供四方面之优化建议:

第一,明确法律知识的主地位。智能司法系统以辅助法律决策为本位,其基础设施建设也应以法律要件知识为本体,而非依赖于纯粹的智能科学技术。现有不少司法人工智能的理论模型或实践应用主张法律与计算技术的交叉,^[59]指出应将法学知识与技术能力置于同等重要之地位,并就此提倡大力培养具有复合学科背景的应用型人才。^[60] 笔者认为该路径虽具有一定的合理性,

[56] 参见纪海龙:《法教义学:力量与弱点》,载《交大法学》2015年第2期,第90—104页。

[57] 参见卜元石:《法教义学的显性化与作为方法的法教义学》,载《南大法学》2020年第1期,第49—80页。

[58] See Katsumi Nitta & Ken Satoh, *AI Applications to the Law Domain in Japan*, 7 *Asian Journal of Law and Society* 471 (2020).

[59] 参见马长山:《AI法律、法律AI及“第三道路”》,载《浙江社会科学》2019年第12期,第4—11页;洪冬英:《司法如何面向“互联网+”与人工智能等技术革新》,载《法学》2018年第11期,第169—180页。

[60] 参见张妮、徐静村:《计算法学:法律与人工智能的交叉研究》,载《现代法学》2019年第6期,第77—90页;高鲁嘉:《人工智能时代我国司法智慧化的机遇、挑战及发展路径》,载《山东大学学报(哲学社会科学版)》2019年第3期,第115—123页。

但也不可避免地因错置主次而忽视了法律知识的本位性：从应然目标的角度而言，智能司法系统的应用意义在于辅助审判人员发现法律、适用法律，纵使在此决策过程中有智能算法与技术参与，其核心要义仍是围绕法律规范所展开的司法活动；而从实然内容的角度观之，法律与技术的衔接或交叉工作相对较少，相比于庞大且复杂的法律规范诠释工作而言，其在基础设施中所占的比例并不足道。质言之，法律与技术的复合知识虽对智能司法发展有刚需性，然其所承担的角色仅是沟通两类知识群体的“桥梁”，基础建设的主要内容依旧是“法律获取”的知识网络构建，即：需要建筑沟通之桥梁，而不需要在桥梁上砌造房屋。在此意义上，智能司法基础设施建设需秉持“让专业的人做专业的事”的理念，使法学理论研究者与法律实践工作者不断完善对法律规范的理解，切不可因技术的交叉性而颠覆了法律知识在其中的本位。

第二，深植法律规范的解释论。法律解释素来是法律职业人员的传统技艺，目前随着国内法律规范的体系化发展，尤其是随着各部门法教义学的形成与分化，此类解释论在各类法律场景下的显要地位自无需多言。然与之同时，部分学者在解释论之外亦形成了诸多相关或相对的法律研究范式，^[61]尽管其或可为智能司法基础设施的建设秩序提供一定思路，但也易使法律解释论本身的指向有所模糊。就本质含义而言，法律解释论的内涵在于对先验的“法律规范”进行先验的“诠释研究”，其外延在于与后验对象“社会事实”及后验方法“实证研究”相区别。^[62]故在构建智能司法基础设施的过程中，各方有必要妥善处理此两项外部关系：其一，明确规范对象与事实对象之间的对立关系，虽然其二者存在等置意义且可使目光在其间流转往复，但毕竟其内部存在着难以逾越的因果鸿沟，这便是基础设施由数据型转向知识型的根本基点；其二，协调诠释方法与实证方法的互补关系，实证立场下的离散数学方法可为法律规范的逻辑推演提供量化的计算手段，而该手段实质上是以数理理性填补解释论的语言模糊性，由此形成的基础设施也是对人工构造要件知识的有力补充。就此综合意义观之，在法律解释论的总体立场下对法律规则进行深入剖析，并辅以实证手段梳理其内部要件之间的逻辑联系，必然是智能司法基础设施建设的优化进路。

第三，整合法律规则的体系性。“体系性”要求作为基础设施的法律规则知识应形成结构化的网络，进而可使智能司法系统有效地理解具体规范的涵义，并明确其在体系中的准确定位。一方面，体系内的各项法律规则本身需以逻辑符号进行表示，以尽可能减少法律规范的不确定性。虽然法律规范之语文表述可能或多或少地显现模糊性，但其欲体现的假定、条件与制裁规则必然是确定的，特别是以前束范式的逻辑符号予以归纳表征时，更应在形式上脱离不确定性的窠臼。因此，体系视阈下的法律要件知识规则应以精准化、精细化为目标，使基础设施之优化建设不断向纵深方向发展。另一方面，体系内的不同法律规则之间也需要进行交叉比较与校验，以避免或消解其中可能存在的相互冲突。显而易见的是，若同一案件事实同时对多个法律规则的适用条件，而在制裁层面却被赋予了相互排斥的法律效果，^[63]则其在法律规则的体系中难免形成冲突之张力。即使是人类司法裁判者，其在面临此类同位法冲突、异位法冲突或是新旧法冲突时，亦会产生两难之困境，更勿论本身不具内源性思维能力的机器智能。故对于智能司法基础设施而言，其建设过程中必然需要消弭体系规则的矛盾情形，为此需要梳理不同法律规范对同一调整对象的相似规定与共性要素，并推究其最大的适用范围及其可能的边界，从

[61] 参见左卫民：《实证研究：中国法学的范式转型》，法律出版社2019年版，第30—31页。

[62] 参见张永健、程金华：《法律实证研究的方法坐标》，载《中国法律评论》2018年第6期，第73—89页。

[63] 见前注[63]，拉伦茨书，第341—343页。

而厘定法律规则的完整逻辑体系。

第四，构建法律解释的共同体。社会原本期望人工智能技术可襄助司法审判人员减少主观裁量因素，进而强化其对法律规则本身的理解与适用，但当前我国司法系统在全面信息化的浪潮下往往急于实现智能应用，以至于采用裁判文书数据型信息的模式而“将法律规范本身排除了”^{〔64〕}。对此应当斧正法律人工智能的定位，其不应机械地使社会生活与法律事实强行依附于数理统计模型；相反，智能化的理想意义应是“倒逼”法律解释之发展，通过不断完善法律规则本身的理解与诠释，为未来智能司法系统的设计与开发提供法律思维的“心智”。由此可见，国内的智能司法基础设施有待法学理论与法律实务界一同填补，其不仅需要进一步开拓法律规范解释的局面，更需要逐步形成法律解释的职业共同体。其中，构建“共同体”的目的即是统一立法、司法与执法等多方主体对法律规则的理解，从人工智能模仿人类主体的角度来看，该诠释过程的成果正是确定了其基础设施的长期记忆。可以清晰地预见，不同的解释主体对同一法律规范之涵义解读可能莫衷一是，然其中势必存在相对合理的一则。在此基础上，通过职业共同体内部的不断议论与说服，最终便可析出相对统一的公认见解，^{〔65〕}若能将其作为特定法律规则的稳定教义，智能司法基础设施的知识网络则不难构成。

进一步地，由该法律解释共同体形成的智能基础设施，其未来的实践意义也不局限于司法领域，更可为立法、执法、法律研究等诸方面应用所共享。首先，立法为基础设施提供了规范内容的基本遵循，并缔造了知识型法律信息的要件组成结构，从根本意义上而论，也唯有立法者有权设计或改变这一知识网络的逻辑构造。其次，司法与执法场景则为其他领域筑造了相互联通的桥梁与渠道：具体而言，两大场景通过运用立法所得智能基础设施，可从实践中发掘既有知识网络的不足之处；同时，针对实践中的争议焦点问题，司法人员与执法人员也可交由立法者与法学研究者共同商讨，从而使法律解释共同体内部再度进行商谈与辨析。最终，立法者将其析出的规则设置为稳定的法律规范，以实现法律知识网络的优化，并促进智能基础设施的更新迭代。但实现此类共享共建机制的前提，应是对知识网络等设施形成“立法—司法—执法”的建设共识，这不仅需要强调与推动各法律活动环节的沟通，亦需要在合作对话的基础上保持职能独立。对此，在未来智能司法发展的进程中，有必要将前述共识机制及各方权限予以明确，以保障法律解释共同体在沟通上的对等性与民主性。

结 语

无论是否以人类社会的主观意愿为转移，信息化浪潮均已为司法染上了“图灵蓝”，拥抱人工智能、构建智慧法院体系不仅是法治现代化的发展图景，也必然是可付诸实践的未来战略。现如今，当人工智能应用在司法场景中弊端迭生时，若欲使之及时走出窘境，则必须对其背后的深层要素做出优化。一方面，人工智能永远无法脱离基础设施而存在，既有的数据型基础设施难以实现法律获取的过程，其需要回溯至法律规范的本源层面，进而以知识型基础设施作为智能司法系统的本体支撑。另一方面，基础设施也永远有赖于人类主体而生成，不论知识型基础设施在当下如何接近完美，其仍无法在法律规范与社会事实的动态演化之外保持静止，此即有赖于人类主体为智能司法系统提供运行支撑。考虑到法律本身存在的语义模糊与规范冲突等问题，为了使人

〔64〕 参见季卫东：《人工智能时代的法律议论》，载《法学研究》2019年第6期，第32—49页。

〔65〕 参见季卫东：《法律议论的社会科学研究新范式》，载《中国法学》2015年第6期，第25—41页。

工智能可准确理解其中之意涵,人类对法律的发现与适用就必然先于机器而抵达。恰如学界所期待的那般,技术的意义在于走近法律而非排除法律,对智能司法基础的建设也将推动人类不断发现法律、诠释法律。故在未来可预见的一段时间内,法律解释共同体内的研究者与职业者有必要逐步深入社会生活事实、逐步梳理各法律规范的准确教义,以此完善作为基础设施的规则体系。事实上,该过程诚然揭示了人工智能的本质——机器的心智并非来源于经验之数据,而是来源于人类的心智本身。

Abstract The current intelligent judicial applications in China generally use data-based information as the infrastructure. Due to the triple dilemmas of source, fuzziness, and isolation in the data itself, it is difficult for the application system to achieve substantial legal decision-making assistance. Based on the corresponding characteristics of human-machine thinking, the theory of legal acquisition provides guidance for it. The process of decision-making thinking of an intelligent judicial application is to discover and apply laws, and the infrastructure as the driving source of artificial intelligence should be transformed from the uncertain legal data to the certain knowledge of legal provisions. At the ontological level, the atomized constitutive requirement is the smallest knowledge unit, which is expressed by the rules of the first-order predicate logic and can be explained by judgment to form a systematic knowledge network. At the operational level, the infrastructure depends on the collaborative construction of legislators, judges, administrators and scholars. It needs to build a community of legal interpretations that communicate with each other, and form a coherent system of legal rules based on the method of legal dogmatics, so as to promote smart judicial development moves towards the “knowledge paradigm”.

Keywords Intelligent Judicial Application, Infrastructure Construction, Legal Acquisition, Requirement Knowledge, Legal Dogmatics, Computation Legal Studies

(责任编辑:宾凯)