误区与正道:法律人工智能算法问题的困境、成因与改进

洪 凌 啸
(四川大学 法学院,成都 610207)

摘要:随着 AlphaGo 在围棋领域的成功,人们开始思考是否能够将以“深度学习”算法为代表的计算统计概率型算法移植至法律领域。目前,法律科技公司往往打着“深度学习”“强化学习”“神经网络”等旗号宣传自身的法律人工智能产品,但其实际效果却往往不佳。在司法实践中真正得到使用的仍然是以“知识图谱”为代表的传统符号型算法。而效果较好的、使用了“深度学习”算法的语音文字转换系统也是一种通用型算法,并非为法律领域量身定制。同时,算法还存在着不透明、不公正、不中立等问题。在这一现象背后有商业、技术、人才三方面原因,法律科技公司囿于经济生存压力,不得不选择目前看来最稳妥的传统符号性算法;在技术方面,法律自身的特点以及法律标签数据缺失、法律数据质量不高、代表性不足等缺陷也使统计计算型算法在短期内尚无用武之地;而法律人工智能领域人才的匮乏更是制约其发展的重要掣肘。未来,需要开发一种符号型与统计概率型算法相结合的、专门针对法律领域的新型算法,同时,需要在对算法进行可视化操作的同时,进行算法警告、算法开源与算法审计。

关键词:法律人工智能;算法;深度学习;强化学习;知识图谱
中图分类号:DF0-059 **文献标识码**:A **文章编号**:1000-5315(2020)01-0058-13

收稿日期:2019-09-06
作者简介:洪凌啸(1989—),男,浙江舟山人,四川大学法学院博士研究生,主要研究方向为司法制度、法律人工智能。

“法律机构和律师们正站在十字路口,将面对未来 20 年间的剧烈变革,其变化程度将超越过去两个世纪的总和。”[①]目前,随着大数据与人工智能技术的不断发展,法律体系逐渐向计算机化、流程化以及自动化发展。在中国,“智慧法院”建设如火如荼,新兴的法律科技公司不断地向法院、检察院以及律所推销研发的法律人工智能产品。许多法律人工智能产品被冠之以“深度学习”“强化学习”“知识图谱”等词汇,在法律人看来,皆属莫测高深的算法词藻。然而,在司法实践中,这些法律人工智能产品似乎并未发挥出其应有的作用,“新瓶装旧酒”现象不断出现,法律人工智能领域并未出现如 AlphaGo 这般革命性的产品。同时,算法在司法、执法以及社会其他领域引发了诸如透明性、公正性等诸多问题。因此,有必要对法律人工智能使用中的算法问题进行相应研究,区分“真”算法与“伪”算法,并反思算法在未来的改进方向。

一 法律人工智能算法的困境

(一)算法的名实分离

人工智能进行“学习”的燃料是数据,“引擎”则是算法。一般来说,算法通过对数据的训练来提炼模型,

①理查德·萨斯坎德《法律人的明天会怎样?——法律职业的未来》,何广越译,北京大学出版社 2015 年版,第 1 页。

进而总结出相应的规律并预测未来。有学者认为,算法的派别可分为符号学派、联结学派、进化学派、贝叶斯学派与类推学派。① 从更宏观的角度而言,算法其实可以分为两种:一种是以逻辑推理为基础的符号算法,另一种是以数据概率为基础的计算算法。前者的典型代表是专家系统,发展至今,在法律人工智能领域的代表是"知识图谱"算法。后者的典型代表则是"深度学习"与"强化学习"。人工智能从其诞生至今,经历过数番波折,专家系统从被吹捧到被摒弃,其根本缺陷是其自身的封闭性,只能根据人类专家事先设置的规则进行推理,因此无法对纷繁复杂的现实社会环境的问题进行回应。而最近的人工智能热潮的出现需归结于AlphaGo 在围棋上击败李世石,而围棋的天量运算量在以前普遍被公认为是不可能被机器所取代的,因此被誉为"人类智慧的王冠"。随着 AlphaGo 击败李世石,并在一年后击败柯洁,人们在震惊之余开始思考AlphaGo 这一谷歌 Deep Mind 开发的"人工智能"强大的原因。谷歌团队指出,"深度学习""强化学习"是AlphaGo 成功的关键。这也使人们尤其是法律人想象是否可以将这种强大的算法移植到法律领域,运用在一些简单法律事务自动化处理的基础上,实现法律裁判的智能化。

需要指出的是,当下还没有一套可以适用于各类案件、各种司法实践场景的万能算法,希冀于用一种全能算法、一种通用模型架构来解决司法场景中的所有问题无疑是一种神话。因此,不同场景下的不同法律人工智能产品所使用的算法各不相同,但比较公认的主流算法是"深度学习""强化学习"与"知识图谱"。

"深度学习"算法本质上是一种统计学技术,其通过多层的神经网络技术对数据进行分析,进而建立算法模型对问题进行预测。神经网络技术只是因其输入节点、隐藏节点和输出节点的网状结构连接类似生物神经元之间的连接而得名,但实际上与生物意义上的神经网络毫无关系。神经网络技术是多层的,计算机中的数学"神经元网络"就是一系列像神经元一样可以接收、评估、传递信息的彼此相连的开关。每个开关就是一个数学方程,方程上携带着多种不同的信息投入,并给它们赋予不同的权重。网络的终端是一个总开关,负责收集前面所有神经元开关的信息并生成预测,作为神经元网络的产出。② "深度学习"的大规模运用改变了从前人工智能发展"专家系统"只解决能够清晰表达的问题,不再过分依赖先验知识与固化逻辑,而开始对未来的结果进行预测。"深度学习"并不是一个全新的算法③,它出现于 1980 年代,是计算人工智能的一种。但由于受算力与数据量的制约,它受到的关注度要远小于其他算法。2012 年,Krizhevsky、Sutskever、Hinton 一系列成果的发布,以及在 Image Net 目标识别挑战赛上取得的成功,让"深度学习"算法再次回到人们的视野。看到"深度学习"算法的前景后,国外学者纷纷跟进④,并在 AlphaGo 战胜李世石后名噪一时。

"强化学习"是一种介乎于"监督学习"与"无监督学习"之间的机器学习方法。监督学习所用的数据是固定的标签,"强化学习"则更进一步,其标签并不固定,但可通过固定规则对训练数据进行约束和间接标注。通过对奖励函数的设定,使"奖励"(reward)与"行动"(action)之间的相互关系强化,"强化学习"算法可以不断通过激励函数得到反馈,对特征点的权重进行更新,不断得到强化与修正。"强化学习"在计算机科学理论上可以适用于包括未知信息领域在内的任何事物与环境。

"知识图谱"是在专家系统的基础上发展起来的,相比专家系统,"知识图谱"更加自动化,可以半自动地实现符号逻辑的编排。但是,"知识图谱"的本质依然是通过符号辅以严密的逻辑推理模拟人类的思维方式。因此,"知识图谱"算法属于符号学派,它模拟的是人脑的推理方式,其针对的对象是规则,其比较类似决策树算法。"知识图谱"使用图作为表示知识的数据结构,以"结点—边—节点"的形式组成知识和事实表示的陈

①佩德罗·多明戈斯《终极算法:机器学习和人工智能如何重塑世界》,黄芳萍译,中信出版社 2017 年版,第 66 页。

②伊恩·艾瑞斯《大数据思维与决策》,宫相真译,人民邮电出版社 2014 年版,第 140 页。

③皮埃罗·斯加鲁菲《智能的本质:人工智能与机器人领域的 64 个大问题》,任莉、张建宇译,人民邮电出版社 2017 年版,第 114 页。

④有关"深度学习"的更多资料与参考文献,可参见:Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM* 60, no. 6(2017): 84-90, Doi: 10.1145/3065386; Baolin Peng, Zhengdong Lu, Hang Li, Kam-Fai Wong, "Towards Neural Network-based Reasoning," erpint arXiv: 1508.05508, 2015, https://arxiv.org/pdf/1508.05508v1.pdf; Dan Cierşan, Ueli Meier, Jürgen Schmidhuber, "Multi-column Deep Neural Networks for Image Classification," 2012 IEEE Conference on Computer Vision and Patter Recognition (CVPR)(Washington DC: IEEE Computer Society, 2012), Doi: 10.1109/CVPR.2012.6248110.

述语句。“知识图谱”最大的作用是降低了结构化知识构建和使用的难度。在司法领域,“知识图谱”一般通过对法律知识网络的构建,帮助法律工作者在线快速检索法律条文与知识。这种可视化的分析与信息检索为自然语言识别与理解提供背景知识库,问答系统基本上主要依赖“知识图谱”算法。通过“知识图谱”算法图的表达,大量数据可被压缩,复杂关系与信息的查询与表达被大量简化,查询速度大大加快。

一般来说,数据库是最常用的数据收集、存储与分析方式,通过数据库,机器可以高效地获取信息。数据库的缺点是当数据量过大时,对复杂关系的运算与多度、跨表查询耗时较多,这对算力不够、计算机硬件设备不先进者是非常不利的。同时,这些算法都有其使用的场景与条件。“深度学习”“强化学习”的前提是充分的标签数据。“深度学习”最适合的领域是对数据进行分类,尤其是对非结构化的大数据集进行处理,即通过神经网络输出结果定义的反向传播,给定输入数据的类别。另外,“深度学习”算法具有一定的通用性,可以将其适用到各个领域上,而不需要特别丰富而专业的知识。从这个角度而言,“深度学习”拓展了人工智能的运用领域。一般来说,当数据量充足时,可使用“深度学习”算法进行模型构建。“强化学习”也需要海量的标签数据与学习样本来训练得出一个原先通过硬编码即可简单学习的普遍规律。而当数据量不足时,如只有几百或几千个数据值时,我们就需要通过人工的方式,如引入图模型来构建一个人类的知识体系,而不是由机器自身形成的模型。“知识图谱”可以帮助机器识别与使用来自不同数据源的数据,主要是依靠图表的方式将数据间复杂、交互的交叉关系表现出来。这在数据源愈发多元化、数据存储格式也各不相同的当下显得格外重要。在“知识图谱”的基础上,有公司甚至开始加入时间维度,以生成事理图谱,可视化、智能化地展示案情的事实与证据情况。

最近几年,法律人工智能界使用最多、宣传最广的算法即是“知识图谱”。“知识图谱”确实具有一定的优势。首先,具有知识性,即该算法可以累积较多的专家知识;其次,具有逻辑性,可以通过固定的符号模拟人类的思维方式进行推理、判断并做出相应的决策;最后,具有透明性,即以符号与推理的方式展示人类的思维过程,其所使用的符号数据与推理过程都是可视可解释的。“知识图谱”用计算机表示与推理的形式将专业领域中的经验知识概括、转化为机器能够识别的符号,并将测试数据与之对比、匹配,最后得出推理预测的结论。

在司法实践中,比较常见的法律人工智能产品有法律检索、文书自动生成、类案推送、语音文字转换等。法律检索系统、裁判文书自动生成系统所使用的算法是“深度学习”、类比推理[①]与支持向量机[②]。类案识别和推送所使用的算法是“深度学习”“强化学习”与“知识图谱”;语音文字转换系统所使用的算法是“深度学习”。

由此可见,在司法实践中的法律人工智能产品首先是一种算法的集合或者混合。尽管在名称的选择上,几乎所有的法律人工智能产品都会强调自身使用了“深度学习”“强化学习”等先进的算法,然而从效果上看,法律人工智能产品的实际效果不一。事实上,效果比较好的法律人工智能产品是以科大讯飞为代表的语音文字转换系统,而类案推送、法律检索等法律人工智能产品并未真正得到运用,即使运用也因用户感受不佳而得不到充分运用。例如,有学者在考察类案推送系统的过程中发现,类案推送在司法一线并未得到广泛的运用与好评,甚至有许多法官反映,类案类判系统对法官办案“帮助不大”“作用很小”。[③]

更进一层讲,尽管类案推送与法律检索等效果不佳的法律人工智能产品背后的算法是“深度学习”“强化学习”与“知识图谱”,但真正起作用的却是“知识图谱”。因此,最新的计算统计概率型算法技术其实并未在法律人工智能产品中得到运用,我们仍然在使用传统的符号型算法。而运用效果较好的语音文字转换系统,虽运用了“深度学习”算法,但需要指出的是,这一算法是专门针对语音文字转换领域的算法。在该领域,这是一种通用技术,既可以适用于法律领域,也可以广泛地适用于翻译、教学等与语音、文字相关的领域。也就

①类比推理也称最近邻算法,即通过对相似度的衡量,归纳、重组、推导出创造性的预测意见。

②支持向量机指的是算法先把每个词语都转化为一个“向量”,即多维度的“量”,再将每个词语进行向量化,即“词嵌入”(Word Embedding)。它针对的对象是实例。

③左卫民《如何通过人工智能实现类案类判》,《中国法律评论》2018年第2期,第26-32页。

是说,这种算法并不是专门为法律领域量身定做的。

总体来说,当下法律人工智能领域的算法存在着严重的名实不符现象。首先,几乎所有的法律人工智能产品都会强调自身的算法具有先进性;其次,司法领域中使用效果不佳的法律人工智能产品背后的算法,是以传统型的符号流派算法为主的,而最新的计算统计概率型算法并未真正被运用;最后,运用效果较好的算法背后是一种"通用型"的算法,目前缺少专门为法律领域专门设计的算法。

(二)算法的透明性缺陷

"深度学习"是一个"端到端"(end-to-end)的黑箱,人类无法获知其做出决策的过程、理由与原因。在法律领域,"深度学习"算法的致命缺陷在于给出一个判决结果不等于给出了判决尺度与判决规则。"深度学习"所给出的 YES or NO、RIGHT or WRONG 的答案无法反映判决的全部内容与法律推理过程。此外,机器如何判断 YES or NO、RIGHT or WRONG 也是一个重大的谜团。如何定义胜诉?支持了全部诉讼请求或者驳回全部诉讼请求算胜诉,那么实际损失 80 万,起诉金额 100 万,最后法院判决支持赔付了 50 万,在诉讼上算输还是算赢呢?在刑事案件中,在 3 到 10 年的量刑幅度间,被告人被判了 5 年算胜诉吗?这些问题都无法用 YES or NO、RIGHT or WRONG 来简单定义。例如威斯康星州诉卢米斯一案〔State v. Loomis, 881 N.W.2 d 749(Wis.2016)〕:2013 年,埃里克·卢米斯(Eric Loomis)因偷窃被枪击者抛弃的汽车而被警察误当作枪击者予以逮捕,并受到与驾车枪击有关的五项刑事指控。鉴于卢米斯存在偷盗和拒捕行为,卢米斯承认了其中两项较轻的指控。卢米斯回答了 COMPAS 犯罪风险评估工具所问的一系列问题,并被 COMPAS 认定再犯可能性是"高风险"。COMPAS 系统是威斯康星州惩戒部门一直使用的,由一家私人持股公司开发的一款风险评估工具:一种基于证据衡量罪犯未来犯罪可能性,并为矫治署提供决策支持的软件。风险评估算法的技术原理是,罪犯先回答一系列问题、问卷或采访。例如 COMPAS 就有 137 个问题的对话系统,这些问题涉及犯罪和个人历史,包括家庭犯罪历史,同时也涉及很多的个人观念和看法,比如个人可信度、对场景善恶的判断等等。COMPAS 的问卷将这些问题分为 15 个维度:当前指控、犯罪历史、不遵守、家族犯罪性、同辈交往、毒品滥用、住所的稳定性、社会环境、教育、职业、空闲与娱乐、社交孤立、犯罪人格、愤怒以及犯罪态度。之后,算法对所有数据进行处理,判定罪犯的再犯罪风险级别。此外,还涉及需求级别,用于对犯罪人的教育和改造。级别分是在同基准群体与其他罪犯比较的基础上得出的,1—4 为低,5—7 为中,8—10 为高。用于比较的基准群体有男女两类共八组:男性监禁/假释、男性监禁、男性缓刑、男性混合,女性监禁/假释、女性监禁、女性缓刑、女性混合。以此来判断罪犯的个人身份信息、成长经历及种族状况。① 法庭在量刑时,参考了 COMPAS 犯罪风险评估以及其他众多因素,将 COMPAS 的犯罪风险评估作为对卢米斯量刑前调查报告(PSI)的一部分,最终判处卢米斯 6 年监禁和 5 年监外执行。之后,卢米斯提起上讼,主张法庭严重依赖 COMPAS 系统进行判案的行为侵犯了其在美国宪法第五、十四条修正案下所享有的正当程序权利,即量刑需要注重个案主义及量刑的准确性,而 COMPAS 的私企性质及商业秘密阻碍了其评估的准确性;并且,COMPAS 系统不当地考虑了性别因素,其评估结果的非准确性使其不能作为判案依据。另一方面,法院根据 COMPAS 系统的预测结果进行判案有程序违法之嫌,不符合个案处理的原则。因为在判审期间,卢米斯并没有接触这个算法的权限。2016 年 7 月,美国威斯康星州最高法院支持了下级法院的裁判,驳回了卢米斯的请求,认为初审法院在量刑时利用犯罪风险评估分数不侵犯被告人的正当程序权利,并且将性别作为参考因素反而提高了犯罪风险评估的准确性。② 即使背后的算法和方法没有向法院和被告人披露,算法输出的信息有着足够的透明度。在准确量刑方面,一方面,在评估犯罪风险时考虑性别正是为了提高准确性;另一方面,COMPAS 使用的公开数据都是被告人提供的,如果有错误,被告人可以"反

①朱体正《人工智能辅助刑事裁判的不确定性风险及其防范——美国威斯康星州诉卢米斯案的启示》,《浙江社会科学》2018 年第 6 期,第 76-85 页。

②Michelle Liu, "Supreme Court passes on crime assessment case," accessed November 12, 2019, https://www.jsonline.com/story/news/crime/2017/06/26/supre-court-refuses-wisconsin-predictive-crime-assessment-case/428240001.按:本文所引外文文献均为笔者自己翻译,下同。

驳、补充和解释”这些信息。COMPAS 系统的风险评估是借助独立的子项和复杂的算法完成的,最终从 1 到 10 的级别评定具有中立性和客观性。但法院同时强调,在使用犯罪风险评估工具前,应当给予法官诸如算法的公开性、有效性、歧视性的提示,提醒法官在量刑时不要过度依赖机器算法,不要过度使用算法。2017 年 6 月,美国最高法院拒绝提审该案,这实际上间接承认了法律人工智能在司法实践中运用的现状,并希望维持现状不变。这同时意味着美国法院在法律人工智能的使用问题上尚未形成共识,需要用时间来消化科技带来的冲击。

司法权作为一种中立、被动的权力,相较立法权缺失了民意基础,相较行政权缺失了强制手段,但承担着纠纷解决的终局权威、实现社会正义的最后堡垒的重任。司法权获得正当性的重要基础在于其理性,而理性需要通过司法裁判的说理机制予以体现。法律判决不断规范、完善的过程是一个不断增加其说理性的过程。一个判决只有能够被解释,才存在被评估、被信赖的空间。也正基于此,我们才可对其进行修正以增加共识,减少司法判决不透明所带来的震荡与风险,并推动相关法律领域立法的进步与完善。算法尤其是“深度学习”算法在其运算过程中的方式往往是人类所无法完全理解的。这种在数据输入与输出之间不透明的状态被形象地称为“黑箱”。也正因为此,我们对机器这种复杂到人类都难以理解的自我学习、自我演进方式,基于数据建立模型并给出答案的能力感到迷惑与担忧。

“以前,人类是所有重要问题的决策者;而今,算法与人类共同扮演这一角色。”①但不管人工智能有多复杂,其实质还是统计科学与计算机科学的结合,依然是数据与代码的排列组合。因此,需要算法的设计者在设计伊始便从算法内部增强其解释性。

(三)算法的公正性与中立性忧思

算法表面上并未依靠暴力来维持与推动,并且在长时间的话语渲染下披上了一层科技化的神秘外衣,树立起不容质疑的隐形权威,如果没有算法专家的帮助,普通群众更是对算法难以“去魅”。但恰恰是这种对算法中立性与公正性的盲目迷信,引发了狂热的数据与算法崇拜的思潮,“数学洗白”(math washing)现象也愈发严重,人们不断让渡出自己对事物的判断权与决定权,而算法则不停地在新的领域开疆拓土,占据话语上的统治地位。这种通过算法所建立起来的新型支配关系,正在演变为一种新兴的权力——算法权力(Algorithmic Power)②,正在培育新的不平等空间。算法在不断自动化地为公众提供现实的答案的同时,也带来了新的问题。

这其中,算法的公正性与中立性问题最为引人注目。算法或者说技术是中立的吗?显然,公众对算法中立存在重大误解。每个人对算法都有着良好的期待,寄希望于算法能够更加客观而无偏见地给出预测与结论,而不受人类主观情感与情绪以及运算能力的影响。事实上,算法也确实在某些方面与程度达到了这一期待。例如,有研究显示,法官在假释与保释环节容易受罪犯外表长相的影响,做出不正确的结论;而算法则不会受这些人类个人情感因素与主观好恶的影响,个案判决间的偏离度更小,也显得更加公正与中立。有鉴于此,在对外宣传上,法律科技公司一直宣扬着自身的高效、自动、中立与公正,但事实真的如此吗?非常遗憾的是,我们或许正在见证技术的另一种偏见。正如威廉·布鲁斯·卡梅隆所指出的——“并非所有能够量化的东西都很重要,并非所有重要的东西都能量化”③,但“个人的无意识被掌握在算法手中”④。法律人工智能行业中的算法是作为一项商业秘密而存在的,外人无从知晓,只有算法的设计者才掌握具体细节。算法中夹杂了太多的商业利益、政治考量与文化偏见。例如卡内基梅隆大学利用一种名为 AdFisher 的广告钓鱼软件,模拟普通用户浏览求职网站的行为。结果发现,由谷歌推送的“年薪 20 万美元的以上职位”男性用户组收到 1852 次推送,女性用户组仅仅收到 318 次。研究者认为,谷歌公司的广告系统已经学会了性别歧视。⑤

---

①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。

②郑戈《算法的法律与法律的算法》,《中国法律评论》2018 年第 2 期,第 66-85 页。

③安德雷斯·韦思岸《大数据和我们:如何更好地从后隐私经济中获益?》,胡小锐、李凯平译,中信出版社 2016 年版,第 177 页。

④瑟格·阿比特博、吉尔·多维克《算法小时代:从数学到生活的历变》,任轶译,人民邮电出版社 2017 年版,第 133 页。

⑤Claire Cain Miller, “When Algorithms Discriminate,” *New York Times*, July 9, 2015.

美国联邦贸易委员会在调查中发现广告商更倾向于将高息贷款信息展示给低收入群体看。[①] 再例如“今日头条”等个性化新闻软件的出现,会让市民只接触迎合他们狭隘偏好的新闻出版物,从而形成“我的日报”(Daily Me)。[②]

在刑事司法领域,这一现象尤为突出。有数据表明,被警察拦截搜身的男性中,黑人或拉丁美洲裔人的比例高达85%以上。[③] 这在某种程度上加大了如未成年饮酒及公共场所抽烟等轻微罪的被发现与放大。一旦这些黑人与拉丁美洲裔人控制不住情绪与警方产生冲突并因此被捕的话,他们就有了犯罪前科。而这些有色人种多集中聚居在一些贫困的社区与街道,由于犯罪前科的出现,该地的历史犯罪率自然就会进一步提高,算法由此会指引警察去此处进行预防型巡逻,这就导致更多的黑人与拉丁美洲裔人被盘查与搜捕,被捕率进一步提高,进而形成了一个恶性循环。有色人种经常居住地的犯罪率居高不下,证明了加强警察巡逻的必要性,警察巡逻造成更多的轻微罪的犯罪率与犯罪前科,算法由此更加倾向于对有色人种进行巡检与拦截搜身,这是一个失真而有害的恶性循环。纽约公民自由联盟2013年的调查数据显示,虽然14至24岁的黑人和拉美裔男性仅占纽约人口4.7%,但警方拦截搜身的对象高达40.6%属这一群人。[④] 马里兰大学一项研究显示,在包含休斯顿的哈里斯郡,相对于被判犯了相同罪行的白人,黑人被检方求处死刑的几率高三倍,西班牙语裔被求处死刑的几率高四倍,而且这种形态并非德州独有;美国公民自由联盟指出,在联邦系统中,黑人得到的刑期比犯类似罪行的白人长约20%,而黑人虽然仅占美国人口13%,但美国在囚犯人高达40%为黑人。[⑤] 而从犯罪类型看,目前算法能够进行预测的犯罪类型往往是街头犯罪、常规犯罪及与人身相关的恶性犯罪,但金融犯罪、欺诈犯罪、白领犯罪与高智商犯罪却不在其列。可以说,算法的精准与高效也是针对穷人的精准与高效,而富人这一群体在刑事司法领域被算法有意无意地忽略了。“未来,富人的事务会由人打理,平民的事情则交由机器。”[⑥]通过算法的不平等、不公正、不中立的统治将变得更为隐密,手段将更为精细、间接与难以察觉。犯罪概率评估系统工具的使用需要考虑公共利益,这其中,该工具赖以存在的算法的合法性与公开性构成了这类工具的合宪性前提。美国调查性新闻机构ProPublica最新的一项实证调查研究表明,COMPAS已对黑人造成了系统性的歧视。COMPAS系统将黑人错误评估为高犯罪风险及罪犯潜在分子的概率几乎是白人的两倍。[⑦] 实证研究显示,被COMPAS系统认定犯罪风险程度相当的黑人与白人,当二者被假释后,白人却更有可能(far more likely)重新犯罪。这就意味着,COMPAS系统将白人认定为低风险的做法是不准确的。甚至有学者认为,COMPAS系统在预测未来犯罪方面的准确性和掷硬币差不多。另外,Tan和Caruana根据COMPAS所描述介绍的指标体系构建了一个模拟COMPAS的模型,同时,他们还设置了一个对照组,即基于现实世界的实际再犯结果创建了另一个模型。通过对实验组及对照组模型的比较,Tan和Caruana根据输出结果与种族、性别各变量之间的关系进行比对后发现,COMPAS确实对黑人存在系统性偏见。[⑧] 另有研究者指出,风险评估算法的具体内容被商业公司的保密协议所保护着,要想获得具体的分析数据、算法与结果是不可能的。[⑨] 此外,更惊人的是,风险评估算法已逐步向“深度学习”算法转变,这就让原本可评估、透明可视的算法变成“黑箱”,“深度学习”会通过运算自己得出相关的结论,但这

①苏令银《透视人工智能背后的“算法歧视”》,《中国社会科学报》2017年10月10日,第5版。

②伊恩·艾瑞斯《大数据思维与决策》,第23页。

③国务院新闻办公室《2010年美国的人权纪录》,《人权报告》2011年第3期,第2-11页。

④凯西·欧尼尔《大数据的傲慢与偏见:一个圈内数学家对演算法霸权的警告与揭发》,许瑞宋译,中国台湾大学出版社2017年版,第41页。

⑤凯西·欧尼尔《大数据的傲慢与偏见:一个圈内数学家对演算法霸权的警告与揭发》,第40页。

⑥《算法密码之凯西·奥尼尔:盲目信仰大数据的时代必须结束》,2018年10月29日访问,https://new.qq.com/omn/20180203/20180203A04R1Z.html.

⑦Jeff Larson, Surya Mattu, Lauren Kirchner and Julia Angwin, “How We Analyzed the COMPAS Recidivism Algorithm,” accesed November 12, 2019, https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

⑧Sarah Tan, Rich Caruana, Giles Hooker, Yin Lou, “Distill-and-Compare: Auditing Black-Box Models Using Transparent Model Distillation,” eprint arxiv: 1710.06169, 2017. doi:10.1145/3278721.3278725.

⑨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,第108-124页。

个过程就连开发的人员也很难解释原因。

归根结底,算法的中立性与公正性取决于数据库的覆盖率与准确度,以及设计者给定的规则的客观性与权威性。如果数据库中数据的代表性与准确性没有问题,而提前设定的规则也未掺入人类的情感偏见,那么这套识别方法在一定程度与范围内是有效的。算法的不公正性有时并不是算法开发者的故意为之。每个人都有特定的身份与社会属性,如种族、性别、家教、学历等,除非每个人都身处无知之幕(a Veil of Ignorance)[①]之后,否则算法所产生的不公正其实是社会不公正的投射。为此,首先,需要区分不公正是算法原因造成的不公正还是社会环境造成的不公正;其次,需要营造一个开放多样的算法竞争市场机制,以避免算法系统性的不公正,即每个人都有自由选择算法,自由选择法律人工智能产品的权利;最后,在司法领域,当事人作为数据主体(Data Subject),有权不接受法律人工智能自动得出的裁判结果,并可要求生产法律人工智能产品的公司提供详细的数据来源、算法模型与输出结果,并对其进行解释。

二 算法"困境":缘何如此

(一)商业原因

"深度学习"算法的基础是海量的标签数据,对法律数据进行标签化处理,虽然不需要特别专业的法律人才,但关键在于数据量太大,因此所需要的人力与资金的支持也是惊人的。可以说,在现阶段,"深度学习"与"强化学习"算法所需的成本与投入是一般的企业无法承受的。这就造成了"强化学习"与"深度学习"算法的话语宣传、学习与掌握与其实际可能之间产生巨大承受或接受鸿沟。就某种程度而言,AlphaGo 在围棋领域成功的宣传意义要大于实际意义,AlphaGo 的胜利是不常见的、偶然的胜利,可能是不具有代表意义的。

对于大公司尤其是中国的大型企业如百度、阿里巴巴、腾讯等头部互联网公司而言,它们当然重视算法,但是它们更希望将其运用于商业如电商领域。或许在法律人眼中,法律行业的经济体量较大,但将其放置于整个中国经济的大环境下,法律行业的经济体量其实并不大。中国头部律所一年的创收额度甚至比不上淘宝、京东等电商企业"双 11"一天的创收数额。此外,法律毕竟是分配蛋糕,而不是制造蛋糕,并不会额外制造经济效应。因此,大企业并没有特别强的动力去研究适用法律领域的算法,更希望将有限的资金与资源投入到能够产生巨额利润的行业领域如医疗、电子商务等。

法律科技公司往往并不是大公司,而是初创公司,因此,法律科技公司往往会回避对深度算法的使用,但为了和人工智能的热点联系在一起,只是在宣传话语上模糊地提到"深度学习",但在实际运用层面上,只是通过法律条文"知识图谱"的构建,将所有相关的法律条文串联起来。

法律科技公司第一需要解决的是生存问题,在面对大众的法律人工智能技术与产品尚未成熟的情况下,法律科技公司既需要在短期利益与长期发展中进行选择,也需要对产品的受众进行区分。由于自然语言技术的制约,法律科技公司无法对日常生活语言进行准确的判别,因此常常会将产品目标对接法律专门机关。同时,许多法律科技公司并不具备如生产出 AlphaGo 的谷歌 Deep Mind 团队的人工智能技术。在"深度学习"算法上,法律科技公司既缺乏大量资金与人力去完成文书标注的基础工作,又缺少 GPU 这种"深度学习"算法必备的硬件条件,因此选择用较为简单的传统算法进行产品构建。

在法律领域,使用"知识图谱"主要切合了当下法律科技公司的实际情况,其主要雇员来自于法院的前法官、检察院的前检察官。在法学专业知识及司法审判技术知识方面,这些人力资源当然是无可挑剔的,也契合"知识图谱"需要对某一特别法甚至是某一特殊罪名细分领域内的法律要素进行识别、选择与构建的要求。这也是法律科技公司引入这些优秀的前法官、前检察官的初始用意。构建"知识图谱"的工程,虽然相较"深度学习"算法的研究要轻松许多,并且在一些简单案件中确实可以起到一定的作用,具有一定的实用性与易用性,但也明显制约了法律人工智能的深度发展。因为当下最简单的道路,在未来很可能是最艰难的道路,并可能将法律人工智能置于悬崖之上。

此外,算法的不透明、不公正、不中立,也在于算法掌握在少数的科技公司与算法工程师的手中。出于商

①约翰·罗尔斯《正义论》,何怀宏、何包钢、廖申白译,中国社会科学出版社 2009 年版,第 105-106 页。

业秘密的考虑,算法以一种隐形的形式存在,并不向外公开。而为了实现利益最大化,算法向某些特别人群倾斜,实践中"大数据杀熟"现象的出现,已经昭示了这一点。

(二)技术原因

算法使用的悖论在于,算法模型越通用,则其虽可容纳下更多的"噪声"①,但其高拟合性自然也降低了预测的精确性。同样的逻辑,当算法模型越个别化,则越只能在特定的场景下使用,越无法容忍数据噪音,但相应的,其精确性会大大提高。人工智能在具体的任务如图像识别、机器翻译上表现出色,可以惊人的速度与准确度完成任务。但在这背后起支撑作用的是独立的算法,即一种任务对应一种算法,算法无法迁移与混杂。当下并没有一个通用的算法可以对所有的法律领域、所有的法律案件类型以及所有的罪名、案由进行概括式、打通式的计算。这就意味着,如果算法尤其是"深度学习"在法律领域未获得突破的前提下,刑事领域将近 470 个罪名、民事领域将近 467 个二级案由就可能需要通过以"知识图谱"的方式一个一个的解决。这也是法律人工智能目前只能在简单案件尤其是个别罪名、案由中运用的原因,这些罪名与案由加起来不会超过 20 个。

1."深度学习"算法的技术缺陷

"深度学习"在围棋领域的大获成功,加上商业宣传话语的加持,让人们对其实际效果产生了误解。需要特别强调的是,人们往往将"深度学习"中的"深"理解为"深度学习"算法可以破解各种深奥的难题,然而事实并非如此。此处的"深"仅就算法技术、架构上的特质而言,指"深度学习"算法具有多个隐藏层,在结构上较"深"。20 世纪 80 年代之后,计算机的算力有了巨大的提升,数据的存储能力有了质的飞跃,能够存储当时看来可称为海量的数据。在此背景下,计算机的存储能力与算力能够支撑其对标签数据进行串联,进而通过高速计算建立模型,寻找事物之间的潜在规律。而这一过程由于与人脑活动中神经元的串联活动相似,故而被称作"神经网络"。

受商业宣传话语的影响,我们将"深度学习"误读为灵丹妙药,似乎可以解决一切难题。然而,事实并非如此。"深度学习"有其擅长的领域,亦有其自身的缺陷。研究者需要做的是,将"深度学习"放置于适合其发挥作用的领域,而尽量回避可能产生错误的环节。"深度学习"的本质是通过大量数据拟合,试图让机器找到特征点。"深度学习"的"神经网络"只认特征点,然后由特征点推算概率。"深度学习"一般分为两步。第一步是将大量训练数据输入到机器中,同时在对应素材上确定标签,之后机器就可以通过 GPU 扫描寻找到标签与数据之间的对应关系,并通过建立模型确立机器所认为的规律。这是机器学习的第一步训练(Train)。第二步则是预测(Predict),即将新的数据输入后,根据之前机器确立的模型,给出相关的预测。

"深度学习",从本质而言,就是一项统计技术。它既然是统计技术,自然有其适用的范围与局限。首先,从数据结构看,"深度学习"擅长在封闭式的数据空间内进行数据分类。特别是当训练集数据的数量足够大,并且与测试集数据在结构与内容上接近甚至相似时,"深度学习"能够出色地完成对数据进行分类的任务。但当训练集数据较为有限且训练集数据与测试集数据大不相同或者出现全新数据时,"深度学习"的泛化(Generalization)能力就开始削弱,甚至无法完成对数据的分类工作。在法律世界中,这就决定了"深度学习"的范围与空间主要限定于案件数量多、案件要素差异不大的案件类型,而疑难复杂的案件则是很难进行"深度学习"的。"深度学习"发挥出色的前提假设,是数据间的差异不大,环境高度稳定。因此,围棋世界因其稳定的规则体系而特别适合"深度学习"的发挥。但在法律世界中,显然不是这套逻辑。其次,从数据层次看,"深度学习"所能够学习、归纳的模型特征还停留在平面的层次上。也就是说,"深度学习"很难学到具有层级关系的数据特征。这就意味着,当数据越难以分类,离背景知识与常识越近时,"深度学习"越无法解决上述问题。一旦缺少大量先验知识的数据,"深度学习"在处理开放性问题上往往是束手无策的。而迄今为止,"深度学习"在将先验知识与背景常识进行归入的工作一直进展不大。再次,从数据数量看,"深度学习""强化学习"的训练量级需达到百万甚至亿的数量级,例如 Deep Mind 在棋牌游戏和 atari 上的研究。最后,

①"噪声"指随机出现而没有相关性的数据,参见:安德雷斯·韦思岸《大数据和我们:如何更好地从后隐私经济中获益?》,第 36 页。

任务与领域是单一的。以图像识别为例,机器在识别鸡、鸭的准确率上强于人类,但也仅限于鸡、鸭。因为"深度学习"是根据鸡、鸭的标签数据进行判断的,如果换一个对象换一个领域,如识别猫、狗,机器则全然不是人类的对手,尽管识别猫与狗要比识别鸡与鸭简单。可以说,深度学习的应用前提是十分严苛的,达不到其中的任何一项条件都会严重影响"深度学习"的效果,无法达到或超越人类的水平。这也是"深度学习"算法一直无法迁移至其他相关领域的重要原因。在现实中,大量工作无法满足上述这四个条件,因此,我们更多的在围棋以及电子游戏中听到"深度学习"算法再获突破的消息。尽管这些消息无疑是鼓舞人心的,但是我们必须看到,"深度学习"算法也仅仅在这些领域表现出色,而实际离我们的现实生活还是比较遥远的。

"深度学习"的前提是需要有大规模的标签数据作为支撑,而现实生活中除非刻意设计或专门投入资金进行攻关,否则很难有高质量、高规模的数据样本出现。"深度学习"的缺点是,当数据量不够大时,有可能会陷入局部极小值的系统次最优解陷阱,即"过拟合"(Overfitting)①。"过拟合"意味着机器将训练数据样本中的某些细节特点做了放大化的处理,将其视作了一般规律。这是相当危险的。特别是面对数据缺乏代表性、结构严重单一、差异过小的局面时,尤显危险。例如,如果训练数据集中无罪判决的数量过小,机器在学习之后就会将无罪判决率低甚至没有无罪判决放大为一般特征,在模型建立完成后,未来即使案件符合无罪判决的条件,机器也会基于之前的数据特征给出有罪之裁判。可见,"深度学习"算法还不能举一反三。②

法律人工智能的建模,需要对对话与上下文理解的大规模数据集。在标注数据时,需要注意前后文、全文甚至行业、常识的背景知识,如果没有大规模的标注数据,法律人工智能是很难取得突破的。而这正是我国法律数据面临的巨大挑战。

首先,我国的法律行业缺少大量有效的标签数据。从整个人工智能界的发展看,获得大量的行业标签数据,将迅速提升该行业人工智能的水平。这种标签数据的量级是千万级的。例如,在图像、语音识别以及机器翻译领域,正是有了前期大量标签数据的积累,才获得了令人瞠目结舌的突破。甚至于当需要在某一领域进行人工智能突破时,会专门组织数据标注的团队,并且该团队是由本领域的专业人才与程序员组成的。以谷歌为例,为实现机器翻译方面人工智能效果的提升,其专门组织了由语言学家与程序员组成的专业团队对数据进行标注。当下法律数据行业面临的重大难题是,对日常生活场景下的自然语词难以用法律语词进行概括与标签化,标准量度不统一、模型目的的不同等因素,则为语素提取、案件要素确定、语义侧重点识别增加了难度。

其次,我国的法律行业缺乏高质量的数据。裁判文书网上所公开的文书的一大弊病是,法官在裁判说理时是以一种"打包说理"的方式进行的。也就是说,对于证据、事实、法律的分析是以一种较为笼统的方式进行阐释的,而不是针对每个证据、每项事实、每条法律进行说理。因此,我们的现有法律数据是笼统的、模糊的,难以进行深度加工与解构。此外,裁判文书的质量以及判决说理的详细程度一直成为广受学界诟病之处③,且不说裁判文书的写作质量,甚至一般的行文措辞都闹出不少笑话。例如,有"裁判文书漏洞迭出","短短一页裁判文书就出现了 7 处错误",甚至还有把性别"女"写成了"吕"的情况。④

最后,中国现有法律行业的数据缺乏代表性,在结构上存在严重的缺陷。⑤ 有些案件类型如醉驾案件上网的数量较为充分,为分析研究提供了充足的资源;但有的案件类型如未成年人犯罪案件、离婚案件、危害国家安全案件、职务犯罪、死刑类案件、无罪案件以及当时当地具有重大影响的案件不上裁判文书网或很少上裁判文书网,这就造成中国法律数据行业中数据的差异性过小。一旦人工智能学习这种在结构上具有重大缺陷的数据集,很可能其归纳、提取的数据模式是局限且具有偏见的,会导致模型的"过拟合"。

---

①Schaffer C, "Overfitting Avoidance as Bias," *Machine Learning*, 10, no.2(1993):153-178.

②伊恩·艾瑞斯《大数据思维与决策》,第 142 页。

③孙海龙《如何提高裁判文书质量》,《中国审判》2013 年第 8 期,第 26-28 页。

④《人民日报刊文评"七错"裁判文书:公平正义需司法公开无死角》,2019 年 11 月 12 日访问,http://news.163.com/17/1122/07/D3R442FN000187VE.html。

⑤左卫民《迈向大数据法律研究》,《法学研究》2018 年第 4 期,第 139-150 页。

2.“知识图谱”算法的技术缺陷

“知识图谱”算法的吊诡之处在于,正是因为“知识图谱”算法在计算推理过程上的透明性与可预测性,让人们觉得其与真正的人工智能相去甚远。因为人们对智能的期待往往是“超越人类”的,而这一判断标准的具体化就是人们需要看不懂、摸不透机器在“想”什么。这让最终会形成一串检索树形图的专家系统让人感觉缺少了真正人工智能的神秘感。

在司法实践中,“知识图谱”算法与司法改革中的要素式审判不谋而合,因此得到了广泛的运用,但“知识图谱”算法的缺陷也很明显。一是知识的获取与表述有相当难度。一方面,人类的经验知识在学习与传授的过程中具有概括性与模糊性,因此难以用准确的符号与规则加以描述与表达,法律事务的理解在转化为清晰的推理逻辑时往往与设想的理想状态有较大差距;另一方面,知识的表述过程是一个浩大且艰巨的过程,要想将专家经验表述清楚,模型的构建需精密且严谨,一旦出现漏洞,整个专家系统的准确性将从根本上崩塌。二是对已有的经验知识要求高。专家系统严重依赖规则推理,其推理前提是专业领域中的经验知识,这就要求这种经验知识首先是正确的,其次是丰富的。如果专家的经验知识在某一问题上尚无定论或者分歧较大,那么就应当慎重使用“知识图谱”算法。因为如果超出了经验知识的范围或经验意见不一致,就很有可能出现算法无法求解、输出错误,甚至“知识图谱”因前后逻辑不一而产生冲突、出现崩溃的风险。三是运用成本高。专家经验的获取需要业内顶尖专家的权威意见,而获取这些信息的成本开销巨大。另外,“知识图谱”算法需要程序员既对某个领域的专业知识十分熟悉,又要熟练掌握编程知识,这就对用人成本提出了极高的要求。而新发展起来的“深度学习”算法则不存在这个问题,“深度学习”算法并不要求程序员熟悉特定领域,只要有专业人员为其提供建模所需的标签数据即可。四是实时性差。“知识图谱”一般适用于数据规模较小的领域,并受制于单一数据源。一旦数据出现异类,则对服务器的负载加重,难以及时给出结论。五是更新迭代能力差。“知识图谱”的本质是专家系统,专家系统的核心是规则推理,而涉及规则推理必然涉及到推理逻辑的固化。因此,“知识图谱”是一种静态而非动态的算法体系,其无法根据更新后的数据自动学习、归纳新的规则,无法对知识库进行迭代,一旦出现新数据、新问题,则需要算法设计师重新进行设置,这严重阻碍了“知识图谱”算法的成长。“知识图谱”这种固化了的逻辑处理专家系统,在真正复杂问题的处理上是束手无策的。实践中的法律问题千差万别、千奇百怪,不可能根据专家系统事先设计好的程序按照机器的意思来发生。尽管“知识图谱”确实具有透明化的优点,但其在面对真实案情时却难以自主学习与实时响应,难以输出多个以上的查询结果,难以具备强大的适应能力和知识获取能力,难以对复杂场景进行智能分析,这也成为人工智能陷入低谷、普遍被认为是辅助手段的重要原因。

(三)人才原因

目前,法律人工智能行业缺少大量法律与计算机科学交叉集合的人才储备。这不仅是法学院的教育体系暂时还无法培养出法律与人工智能交叉学科的人才,更在于在吸引现有的人工智能人才方面,法律行业的吸引力也远远不及大型科技公司。大型科技公司通过对人工智能类创业公司的并购,成功获得了大量优秀的人工智能领域人才,而谷歌、脸书、阿里巴巴、亚马逊更是几乎囊括了这个行业所有的精英团队。可以说,人工智能领域真正的人才库规模其实并不大,因此一旦科技巨头公司完成了人才的搜罗工作之后,法律行业如果没有极其吸引人才的薪资待遇与发展机会,是很难将人工智能人才拉到法律行业里来。遗憾的是,法律行业的普遍薪酬根本无法撼动与挑战互联网科技公司。一个悖论是,鉴于法律领域文本的复杂性,如果程序员能够对法律文本设计出精准的算法,并达到一定法律司法文件分析要求时,他完全有技术能力去其他领域研究与工作。因为法律人工智能行业相较于图像、电子商务等领域,后者的技术门槛更低,市场则更宽广,个人收益也更高。

三　法律人工智能算法的改进

未来,需要建设一个算法实验平台,对算法进行实验、拆分、组合,寻找出不同司法场景下最适合的算法体系,在这个体系中,不是一种或几种算法,而是多种算法的灵活搭配与组合,是一整套算法的系统与架构。

(一)建立符号处理和计算统计混合模型

单一的算法已无法满足法律人工智能的发展需要,应当将以专家系统为代表的符号算法与以"深度学习"为代表的统计算法结合起来。以符号表征系统为本质特征的专家系统已被证明在运行时是十分脆弱的,很大原因是专家系统所处的年代数据与计算机的计算能力比今天要弱得太多。而算法模型的组合(ensemble)可以将各种算法的优点集中起来,从而大幅降低算法的不确定性,虽然还是会出现一定的偏见(bias)。在近年来的 netflix 算法比赛中,第一名及优胜的队伍均使用了算法模型的组合,有的甚至将 100 个以上的算法模型通过叠加高层的方式组合在一起。业内的一个共识是,模型组合是未来的趋势。

如今,一个可行且最新的人工智能科研方向是,将在感知分类领域有着惊人优势的"深度学习"、连接主义、神经网络算法与传统的推理和抽象符号逻辑系统的符号主义、专家系统和规则系统算法结合,既发挥"深度学习"算法在感知输入领域的优势,又发挥专家系统算法在抽象领域分析的优点。目前,这一方向已经有一些尝试性的研究在整合两种算法的讨论上获得了一定的突破。例如 2016 年 Gravesetal 的可微神经计算机方法,Bošnjak、Rocktäschel、Naradowsky 与 Riedel 的可微解释器规划方法,Neelakantan、Le、Abadi、McCallum 和 Amodei 的基于离散运算的神经编程方法。①

另外,相较于无法确定归纳偏置即偏见而一直饱受质疑的黑箱性"深度学习"算法,贝叶斯统计算法可以通过计算归纳偏置确定为有用的算法。贝叶斯网络可以挖掘隐藏的传播节点及其之间的隐含关系,并且可预测隐藏节点后的下一层节点,这是"深度学习"算法所无法做到的。因为如果单纯依赖历史数据,必将会使得通过历史数据训练的模型无法摆脱过去的阴影。因此,为了避免陷入历史数据的陷阱,就需要在历史数据之外加入随机性,而这正是贝叶斯统计算法所擅长的。而级联随机森林算法(Cascade Random Forest)可以模拟法官判案决策逻辑。

未来,可以对司法实践中裁判经验较为成熟的类型案件搭建"知识图谱"。例如,对刑事案件,可以从定罪与量刑要素、证据标准、程序流程等方面制定"知识图谱",对不同类型案件的不同要素进行要素、标准、规则的识别与界定。与此同时,在司法数据沉积累积的基础上使用"深度学习"算法预测与判断案件结果。

(二)对算法进行可视化改进

随着人工智能的不断发展与深入,人们对算法黑箱问题的重视程度也愈发强烈。出于技术以及企业商业化的考量,人工智能所做出的决策的算法过程是不被公开的。未来,通过政府、行业与企业的共同努力,随着对算法的透明性与可解释性做出承诺的公司越来越多,那些拒绝做出承诺的公司将从市场上被逐步淘汰。最新的算法研究已表明,至少在累犯预测方面,由杜克大学计算机科学及电气和计算机工程系副教授 Cynthia Rudin 所设计的具有可解释性的算法模型的准确性与 COMPAS 等黑箱算法的准确性是不相上下的。

算法应当具有人本主义,人在算法的审核中必须起到不可替代的作用。正如凯西所指出的,想要"规管算法,驯服算法",就要让"算法指出可疑之处,由人类去完成最后的核查","它们(算法)的运作必须是透明的:我们必须知道它们接受哪些数据输入,产生什么结果,而且它们必须接受稽查"。② 算法的透明性与可解释性可以根据公共事务的程度进行一定的区分。企业完全的市场商业行为可以采用黑箱算法,但是涉及到社会公共事务尤其是刑事司法、政务公开以及医疗、养老、教育等核心高敏感的公共事务时就必须提高算法公开性、透明性及可解释性的等级与程度,应使用经过公共审计、测试与审查的算法系统,并遵守相关的数据、算法与输出结果的记录与问责程序,以避免引起严重的正当程序问题。而由市场企业主体提供的高度不透明、不公开的黑箱算法、企业内部算法以及未经审计验证、审核、测试的新算法则不能适用在这些领域。

法律人工智能系统在设计时即应当增加可解释的模块。从算法的角度而言,"深度学习"虽然在预测方面有较大的优势,但在可解释方面却偏弱。而"知识图谱"算法虽然无法很好地在疑难案件的预测方面给出案件的答案,但有透明性的优势。因此,需要将"深度学习"算法与"知识图谱"算法结合起来。此外,加州伯

① Arvind, Neelakantan, et al, "Learning a Natural Language Interface with Neural Programmer," *published as a conference paper at* 1*CLR* 2016, arXiv:1611.08945.

②《算法密码之凯西·奥尼尔:盲目信仰大数据的时代必须结束》,2018 年 10 月 29 日访问,https://new.qq.com/omn/20180203/20180203A04R1Z.html。

克利大学的学者认为,可以通过交互式诊断的方式分析人工智能模块的记录情况,并忠实重现特定决策结果做出的计算过程与该过程的执行情况,并辅助确定何种输入特征导致了该特定结果。① 当然,如果这种可解释性的模块过多,可能会降低整体算法的运算速度与效率,严重的甚至还会影响算法在运算结果上的精确度。

从政府角度而言,如果政府能够在经济支持上更倾向于帮助更具有解释性与透明度的算法,法官在审理案件进行判决时拒绝采用无解释性且不透明的算法,那么无疑会起到正向的指引作用,鼓励企业对算法的计算过程与所做出的具体决策进行详细释明。社会公众还应当具有对算法所做出的决策提出质疑并获得救济的权利。

(三)建立算法警告、算法开源与算法审计制度

在法律领域使用算法进行相关决策性活动时,必须附随法院对法官的警告。第一,软件具有不透明性,其商业用途的性质阻止了其对风险分数计算过程的披露;第二,风险分数不能识别特定高风险个体;第三,风险评估基于全国样本,没有针对特定地区的居民进行交叉验证;第四,风险分数引发了将少数种族或特定人群的犯罪人评估为高犯罪风险的问题;第五,风险评估算法主要用于帮助监狱部门的量刑后决定,如犯罪人教育和改造,不得将风险分数用来"决定是否监禁罪犯"或者"决定量刑的严重性";第六,必须持续维护、监测、调整算法以确保其准确性,包括可能在法庭上对算法进行交叉询问。

但仅仅通过警告的手段,在法律人工智能的使用问题上踩刹车是远远不够的。我们还需要算法开源(Open Source)与算法审计(Algorithmic Audit)来怀疑以 COMPAS 为代表的犯罪评估算法的准确性和有效性,对犯罪风险评估作出限制。算法开源指的是,通过开源实现算法透明性,包括被告人在内的任何人可以调查、审查算法。国际社会应当倡议在刑事司法、医疗、福利、教育等核心公共机构禁止使用"黑箱"人工智能与算法系统。算法审计指的是,需要中立的第三方在个案中或者一般地对算法进行审查,而不是由算法的提供者对算法进行准确性和有效性的审查。第三方审查可以确保算法准确性、有效性以及算法得到合理的使用,而算法提供者自身的审查出于利益相关性,显然很难保证中立性与公正性。

但从现实的角度来看,算法开源面临着诸多困难,或许并不是当下最优的选择。因为这首先涉及到企业的商业秘密;其次,即使是企业内部,也无法对其算法得出结果的过程做出充分合理的解释。法律人工智能产品在发布之前需经过严格的检验以确保其不会因实验数据、算法及人类设定的训练规则而产生或放大偏见与错误。训练数据应被确保已清除了诸如性别、年龄与种族在内的已知的偏见。并且,实验的方法、数据与最终结果以及所建立的模型、所使用的算法、所做出的决策应被客观记录且能被查询与使用,方便未来出现问题时可随时进行审查。使用的训练数据的来源及内容应当能够被如实描述。在此过程中,可以建立实验组与对照组进行对比,经内部模拟检查新算法是否可能会有算法歧视与黑箱方面的问题。这种严格的测试是必须的,一旦算法在司法实践中实际运行开来,无偏见的算法会为弥补社会中尤其是刑事司法、警务活动根深蒂固的偏见起到重大的推动作用,形成一个良性的循环,加速社会共同体的构建与形成。法律人工智能产品发布后,企业、政府与科研机构应当共同对其在实践中的运行状况进行监督与持续检测,检测的方法、数据与结果也同样应被公开,供公众查询与了解。在此过程中,既需要对法律人工智能训练所使用的训练数据集进行跟踪与测评,也需要定期对法律人工智能所使用的算法与规则进行反思。在法律人工智能产品的整个开发过程中,需要政府、企业与科研机构共同制定一个能够理解、检测、缓解、超越算法偏见、歧视与狭隘的标准体系。算法的偏见与歧视问题是社会、文化领域中偏见与歧视的映射,这是长期且结构性的问题。特别是在刑事司法领域,歧视问题有其自身的历史遗留问题。因此,妄图一次性地解决算法歧视问题,是不现实的,也过分简化了社会系统的复杂性。法律人工智能行业应努力将法律学者、心理学者、社会学者以及计算机科学与工程学的专家整合一处,赋予他们决策权,通过社会各领域人士的共同努力与跨学科合作研究,

①Ion Stoica, etc., A Berkeley View of Systems Challenges for AI, "Technical Report No. UCB/EECS-2017-159", Accessed November 12, 2019, https://www2.eecs.berkeley.edu/Pubs/TechRpts/2017/EECS-2017-159.html.

借鉴各领域的专业知识,寻找潜在的歧视问题。在此基础上,公开、严谨地制定算法公平性审查标准,并定期更新修订,确保算法检测标准体系的规范化与持续性。此外,算法公平性的监督与问责机制应是强有力的,即使企业无法详细解释算法产生决策的过程,其也应当对算法决策所产生的后果负责。唯有如此,方可促使企业在设计、检测、运用算法时更加谨慎与小心。政府、行业与企业均应促成法律人工智能算法的公开性、透明性与可解释性,增进算法的可信度。政府应意识到算法黑箱性所可能带来的偏见与歧视风险,可从知识产权法律保护、法律法规硬性要求、适当的法律责任分配以及市场监管等方面鼓励、支持企业进行算法开源,行业应当制定算法公开性、透明性与可解释性的相关伦理与规范。

综上,算法问题是法律人工智能中的核心问题,但目前法律人工智能市场上存在着大量"拉虎皮作大旗"的现象,法律人工智能产品中的算法往往有名无实。而在学界,限于学科背景与学科界限,学者对算法尤其是法律领域可适用的算法的研究,无法从计算机科学与统计学等交叉学科方面进行切入与探讨。未来,需要从算法本身及技术方面讨论算法在使用过程中出现的法律问题,并做相关改进与发展,以期构建一种负责任的算法(Accountable Algorithm)、理性的算法(Reasonable Algorithm)与公正的算法(Equitable Algorithm)。

Misunderstandings and Correct Ways: the Dilemma, Causes and Improvements of Algorithm Problems in Legal Artificial Intelligence

HONG Ling-xiao

(Law School, Sichuan University, Chengdu, Sichuan 610207, China)

Abstract: With the success of AlphaGo in the field of Go, people have begun to think about whether it is possible to transplant the statistical probabilistic calculation algorithms represented by "deep learning" algorithms into the legal field. At present, legal technology companies often promote their legal artificial intelligence products in the name of "deep learning", "reinforcement learning", and "neural network", which do not work well in practice. What is really used in judicial practice is still the traditional symbolic algorithm represented by the "knowledge map". The voice-to-text conversion system, which uses the "deep learning" algorithm, is also a general-purpose algorithm, not tailor-made for the legal field. It has problems such as opacity, injustice, and lack of neutrality. The three reasons behind this phenomenon are business, technology and talents. Legal technology companies have to choose the traditional symbolic algorithms that seem to be the most secure at present due to economic reason. Technically, the characteristics of the law itself also make statistical probabilistic algorithm useless in the short term. The lack of talents in legal artificial intelligence is a major constraint to its development. In the future, it is necessary to develop a new type of algorithm specifically for the legal field that combines symbolic and statistical probabilistic algorithms. Meanwhile, it is necessary to perform algorithm warnings, algorithm open source, and algorithm audits while visualizing the algorithms.

Key words: legal artificial intelligence; algorithm; deep learning; reinforcement learning; knowledge map

〔责任编辑:苏雪梅〕