

法律与人工智能的法哲学思考

——以大数据深度学习为考察重点

吴旭阳*

内容摘要:大数据深度学习模式是近年人工智能获得飞速发展的重要模式,其在法律科技中也有相关的应用,并引发了部分法律人的担忧或欢迎。对这种模式在法律中的若干应用进行分析,发现其在现有应用中已经具有大数据的优势,获得不少认可;但也存在智能水平较低、事实认定方面的能力较弱等问题进行法哲学探讨,发现该模式存在对于社会信息收集不广、也不能符合司法能动主义创新要求的缺陷。同时,该模式不能够进行证据的充分质证、事实认定方面的认知能力不足;并在行为的正当性、决策中的潜意识或者非理性、综合性和未来发展、公法的社会性大问题等方面的考量不足。该模式能够协助法律人进行工作与研究,但不可能取代人类法律人。

关键词:法律 人工智能 大数据深度学习 事实认定 行为正当性

一、人工智能的发展及其在法律中的运用

(一)人工智能的发展

自从阿尔法狗战胜围棋界诸多世界高手之后,人类恍然发现人工智能(以下简称AI)在某一方面的智识水平已经超越了人类。以往“深蓝”在国际象棋领域战胜人类并未引发人们的关注、思考或者忧虑;但现在大家显然发现,在不久的将来,AI可能会在很多领域达到人类的水平,辅助人类进行部分工作;甚至超越人类,在部分工作中取代人类。有人甚至会担忧未来高智慧的AI对人类的统治。因此,普通人对于AI迅速由漠视或忽视,改为重视、欢喜或者担忧;人类对于人工智能的关注在近两年呈现爆发式增长。

而在法律领域,法律人们除了关心普通大众所关注的问题之外,还更关心AI在法律中能够起到什么作用? AI能否成为法律人的助手? 还是成为法律人的威胁? 这些都是部分法律人相当关注和思考的问题。这其中,对法律人而言最好的模式当然是AI能够成为法律人良好的助手,而且还不会威胁和取代法律人的职业。事情当真如此吗? 我们可以作一些研究或者思索。

(二)人工智能在法律中的应用及其不足

乘着AI的东风,国内外也出现了不少法律人工智能的创业项目。现有的AI技术和创业团队,进行

*中国政法大学证据科学研究院博士后,厦门大学法学院副教授。

了法律人工智能的研究和创业,并在实践中进行了若干试用。他们力图让法律或者司法裁判、法律教学领域在不久的将来像围棋领域一样,使AI成为重要的角色,部分甚或全部取代人类裁判者、法学教授的作用。

笔者经过文献查阅,以及对若干法律人工智能项目的现场调查、观摩,发现已有的法律人工智能创业项目多数有以下特点:

1.大量地以裁判文书或者法律资料为基本的数据库进行大数据的深度学习。最近几年,AI研究获得巨大的突破,其科学基础在于使用了机器的“大数据深度学习”的技术。现有的法律人工智能项目,也多以该技术模式为主。在中国,诸多法律人工智能项目以中国裁判文书网的数据库为大数据进行AI的深度学习训练。^[1]西方比较有名的法律人工智能机器人ROSS平台的研究2014年始于多伦多大学,先是进行10个月的破产法学习后,它获得了Baker & Hostetler律师事务所的工作。IBM将继续教授ROSS不同领域的法律,以便其在多个法律领域进行工作。^[2]

截至2017年6月,中国裁判文书网的上网文书有近三千万份,以近四年的为主;而其内网运行的亦有六千万份。相关数据有较大的价值,对于司法管理、司法公开、法律教学科研以及相关的法律创业项目而言,均是一座巨大的宝矿。但是该数据库里的文书,对于法律人工智能的学习而言有以下缺点:(1)部分法官因为裁判文书上网公开的原因,也因案件越来越多而精力越来越不够的原因,将说理部分尽量压缩。(2)部分的内部规定、会议纪要等,虽然不会在裁判文书上体现,但却是法官们作出裁判的重要依据。(3)相对判决书而言,案卷是重要的部分,尤其是案件审理的重要内容不少均在内卷之中。但是案卷并不上网,其没有信息化或者电子化,则让研究法律人工智能的部分内容显得有“隔靴搔痒”之感。

在大数据深度学习的模式之外,还有其他的模式,如部分法律人工智能采用类型案件的关键“争点”或者“关键点”的规则梳理和设定。例如江苏省检察院系统采用的“案管机器人”系统,制定了六百多个程序、证据、事实等方面的规则,设定了一千多个对比点。然后可以根据提出的问题进行反馈、修正。但是此类机器人不具有自主学习能力,是否属于真正意义上的人工智能尚有待商榷。

2.在事实认定方面的能力较弱。现有的法律人工智能主要工作流程在于“要素性事实”→“要件”→“裁判结果”,而“要素性事实”则是以“证据性事实”和“推断性事实”为前提。以大数据深度学习见长的法律人工智能,现阶段在事实认定方面是个较大的缺失,很难实现由“证据性事实”向“推断性事实”的推论,再实现由“推断性事实”到“要素性事实”的推论转化。部分法律人工智能项目在事实认定方面的处理,以当事人个人的陈述为主;在一些标准化较强的案件类型中(如交通事故、劳动争议、医疗伤害类型案件方面),系统对当事人在此类法律领域的各个主要方面进行选项设计。当事人在系统界面上对各类事实认定问题进行选择,以协助法律人工智能进行事实认定。然后,在这个事实认定的基础上进行法律规则的推理,最后得出结论。这种事实认定并未经过庭审的质证,双方也没有证据交换,其准确性相对较低(当然,法官或者陪审员进行的事实认定,也会存在意见不一致)。

3.在当下,法律人工智能的进步程度比较低,其在案件的事实认定、定性等关键环节上不能完全、完美地运行;其水平仅仅像是法学院大一、大二学生的水平。案件判断和分析的正确性(或者说是精确性)不高。其大体上不能作为法官助手,而只能作为法官助手在收集资料时的参考。当然,个别法律人工智能产品已经开始具有初步的应用价值,能够初步提出法律意见,并具有一定的正确性;但是这仅仅是在部分比较容易规范化的法律领域。

从以上梳理可以看出,在国家公权力机关所提供的服务中,最高人民法院因为建立了中国裁判文书网,为社会提供了大数据的支撑服务,从而能够支持在裁判文书大数据基础上的机器深度学习;

[1]笔者多次参观走访国内相关的法律人工智能创业项目,并为部分项目提供指导与协助运行。

[2]IBM人工智能进入法律行业:推世界首位AI律师ROSS,载中国机器人网,<http://www.robot-china.com/news/201605/17/32941.html>,2017年7月18日。

有企业也在此基础上一步步“培养”出法律人工智能项目。与其他国家机关相比,司法机关的公开性、统一性比较好(统一的大数据库),为相关创业项目提供了基本的制度和数据支持。但是,以此为基础所进行的法律人工智能研究开发项目,仅仅处于起步阶段,依旧存在诸多问题,需要采用相关的技术进一步优化和升级,也需要进行哲理上的探讨。

二、大数据深度学习模式的优劣

诚如阿尔法狗模式那样,以大数据深度学习的AI具有较强的视野和经验,在某一方面甚至能够远超一般人和专家。在法律领域,采用了这种模式的法律人工智能,也能够依靠法律数据库的大数据进行深度学习,从而实现决策的优化。那么,法律人工智能能否像法官或者律师那样办案?它们能否作为合格的法律助手,甚至在很多工作岗位上起到法官、律师的作用,乃至取代法官和律师?在当下的法律人工智能项目中,这些应该还不能成为可能。但是今后随着AI技术的发展,在大数据的内容比较充分的情况下,其能否通过深度学习而达到、取代法官和律师的作用呢?AI的大数据深度学习的基本模式,优势已经被大家所广泛获知,而其在另一方面也可能存在缺陷。笔者从法律人工智能的视角进行优劣的探讨。

(一)法律人工智能大数据深度学习模式的优势

以大数据为基础的深度学习,能够为法律人工智能提供广博的信息数据基础。当下法律人工智能的主要模式是大数据的深度学习模式,该模式基于庞大的法律数据库。以当下中国为例,就是以数千万份法律裁判文书为基础的中国裁判文书网。通过对此海量裁判文书的深度学习,相信法律人工智能在不远的将来就会有较高的法律知识。这种数千万份裁判文书的海量数据,能够提供较好的决策选择基础,其所提取的裁判决策,也往往是在这些文书中较优甚至是最优的决策,具有较好地解释力和说服力。

以海量数据为基础,是形成较优法律制度的重要基础。我们常言,古代的罗马法是最伟大的法律制度体系,其影响到今日还十分深远;其作为简单商品经济下最完备的法律制度,“以致一切后来的法律都不能对它做任何实质性的修改”(恩格斯语)。从某种角度看,罗马法的这种性质是最大样本的制度筛选结果。也就是说,罗马法从其发展的公元前5世纪开始(就仅以第一部成文法《十二表法》起算)到《国法大全》完稿的公元6世纪,其核心繁荣时期的四五百年时间内,罗马帝国至少五六千万人口,地跨亚、非、拉三大洲;在此广大的时间、人数范围和地域差别内的纠纷解决方案,经过如此大数量的博弈、测试和筛选。这种大数据创新、优选下的法律规则,因其支撑的数据量够大,其适应性自然也比较强,对后世的影响是全球性的和历史性的。因此,现有裁判文书网的大数据已经有近三千万份(内网运行约六千万份),不管存在什么样的不足或者局限,其也有足够的数据量能够培养较为高级的法律人工智能。更何况在此后的一二十年内,裁判文书的数据量至少达到数亿级别的量级;其数据量应该足以支撑更为高级的法律人工智能。

虽然当下的法律人工智能的水平不高,但是以发展的目光看,随着AI技术的提高,以大数据学习为基础的AI存在较大的发展空间;其接近于人类法律人的水平,应该是指日可待的。

(二)该模式下法律人工智能的缺陷

大数据深度学习模式的法律人工智能,其重要优势在于海量的大数据,但是其重要的一个缺点,也是在于大数据;也就是说,法律人工智能学习基础的这种大数据的覆盖面如何覆盖的问题。

第一,法学是一门实践性很强的学科,同时也是牵涉面很广的社会科学;司法决策的过程较长,涉及生活的方方面面。那么,是否能够将世界上大多数有用的数据电子化,或者被数据库所采用?或者人类自身的数据,人类社会交往的各种数据被电子系统所采集,并被归入数据库以供AI进行学习?

从现有的信息采集体系看,基本上是不可能的。尤其是人与人之间的当面交流、行动与合作,在现有的技术水平下基本上不会纳入电脑、手机的信息采集系统中,不会被邮件、互联网购物、互联网社交体系的数据库所采集。

如果从这个角度看,法律人工智能可否有机会理解人类社会的各种行为?其大数据的学习模式就类似于从大学象牙塔毕业出来的法科大学生,带着丰厚的法学知识,却对社会认知不够,在一些案件中就很可能作出比较符合规则或逻辑却不能够为社会所接受的裁判。因此,从这个角度看,法律人工智能不能够作出比较精细、精确的裁判,其更可能仅仅是作为人类司法裁判者的助手。

当然,法律调整的是重要的社会关系,而不是调整所有的社会行为或社会关系。法律是有成本的调整模式,因而只是调整足够重要的社会关系;而在这些社会关系中,法律也不是对所有相关行为进行调整,而往往只是在纠纷发生后被动地进行调整。法律对于纠纷的解决也不强求所有的数据,所以法律人工智能也不需要所有的人类行为数据。但从总体而言,在相当长的一段时间内,由于法律人工智能所理解的人类社会是不周全的,毕竟大量的人类个体信息和社会交往信息在很长时间内并不能上传到网络或者数据库。因此,其不能精确理解这些社会性的规则和纠纷解决方式(当然,在通讯网络的5G或者6G之后,力图实现“万物万联”,其可行性有待于观察)。

第二,从司法能动或者创新的角度看,“世界上没有同样形状叶子”,其实世界上也没有同样的案件。那么我们要问:每一个新案件是不是一类新型案件?或者我们甚至还可以问,在“司法能动主义”盛行的今日,有创意、有想法的法官能不能在普通案件中挖掘新意,创制新的判例?或者在判决书中阐述新的观点,从而发展新的法律学说或者新的权利保护类型、模式?即便是纯粹法学派的代表凯尔森,也意识到“一般规范因司法判决的个别化,始终是对那些尚未由一般规范所决定而且也不能完全由它所决定的因素的决定。所以,在判决内容永不能由既存实体法规范所完全决定这一意义上,法官也始终是一个立法者”。^[3]所以他指出:“立法者,即由宪法授权创造一般法律规范的机关,认识到他所制定的一般规范在某些场合下可能导致不公正的或不公平的结果,因为立法者不可能预见到所有可能发生的具体情况。他因此就授权适用法律机关在适用立法者所创造的一般规范会有一个不能令人满意的结果时,就不适用立法者所创造的一般规范而创造一个新规范。”^[4]

因此,如若法律人工智能的学习模式仅仅是当下的深度学习模式,则其往往只能局限于旧有的知识或者规则,在新的问题上不能很好地以规则的目的或者利益进行解释的扩张或者限缩,从而让规则更好地适应社会不断发展的需要。而不同的社会关系或者犯罪手法层出不穷,先可能是“道高一尺,魔高一丈”;然后法律填补了空白或者漏洞,就出现了“魔高一尺,道高一丈”;然后再反复循环。当代社会出现各类新的法律问题,这些问题及其解决或者规则,在原有深度学习的数据库中没有,则基于此的法律人工智能也不能作出裁判或者不能作出较好裁判。

回溯人类的法律发展史,早在人类最早的法学流派——罗马法中的两大法学派出现了相关的争议,就能很好地说明了这一点。在罗马法兴盛的公元1—3世纪,出现了两大法学派:萨宾派和普洛克鲁斯派。两者差异的其中一个重要方面就是法律解释是否严格按照字面含义进行解释;这就会在实践中产生不同的结果。例如一个典型案例是当时两派关于往牲口身上披红色的衣服使其受到惊吓而跑掉是否是一种盗窃行为的争议。还如关于杀人的解释,拉贝奥(普洛克鲁斯派)在处理杀人案件时(在阿奎利亚法中),会举出“杀”这个词的古代涵义,即使用武器运用暴力杀害一个人。在一个案件中,助产士给女奴一包毒药让其食用致其死亡,拉贝奥则认为法律规定之诉不可用,裁判官应当颁发一个新的特殊诉讼程式。而萨宾则宁可在原有法律体制内提供救济。在这些新的案件问题上,裁判官是否应当颁给一个新的特殊的诉讼程式?还是仅仅局限于原有字面含义?或者对原有字面含义进行

[3][奥]凯尔森:《法与国家的一般理论》,沈宗灵译,中国大百科全书出版社1996年版,第165页。

[4]同上书,第167页。