

【文章编号】1002—6274(2021)01—148—13

# 智能司法中的伦理问题及其应对<sup>\*</sup>

罗洪洋 李相龙

(中南财经政法大学法学院,湖北 武汉 430073)

【内容摘要】人工智能使用于司法涉及到突出的伦理问题,智能司法与司法伦理既有紧密关系又有紧张关系。司法公正的一个很重要因素是证据的获取,证据的获取与知识智慧有很大的关联性,人工智能使用于司法在证据获取等方面具有相对于人工的以下优势:准确性、重复性、快速性、抉择性、无私性。智能司法与司法伦理的紧张关系是指智能司法可能损害司法公正:算法歧视对司法伦理的冲击、改变司法公正的性质和法官主导地位、大数据和机器人出错、人工智能本身没有价值判断。对紧张关系下的司法伦理恶性风险需要未雨绸缪,作出预案:认清人工智能司法的伦理缺位,加强伦理审查;控制对人工智能司法的期望值,坚固法官于司法系统的主导地位;规制算法歧视,保护当事人的人格权。

【关键词】智能司法 司法伦理 法官主导 规制算法歧视

【中图分类号】DF 0-053 【文献标识码】A

人工智能是开发、模拟、利用人的智能并最大程度延伸、扩展和提升人的智能的理论和方法,其标志性成果是具有超强信息处理和加工能力、语言和图像识别能力、数据比对和计算能力的电子机器。斯坦福大学尼尔斯·尼尔森的定义,人工智能是指能够在复杂环境下进行感知、推理、学习、沟通等智能化行为的工具或技术,其核心内容包括两个方面,一是令机器做人类所能做之事、甚至比人类做得更好,二是令机器理解由人类或机器甚至其他动物做出的智能化行为。<sup>[1]P1-2</sup>本文的人工智能定义较为宽泛,既包括智能化的技术,又包括智能化的工具,如机器人。

伦理问题是指对规范人与人、人与自然、人与社会关系之道理与准则的研究。伦理问题的研究是明晰道德意义上的善与恶、是与非,主体于环境之中所担负的道德责任与义务。伦理可看作道德哲学对行为的思考,通过对行为目的、原则规范、价值取向的研究形成普遍适用的行为标准,这种符合道德标准的行为规范,对人与人、人与社会、人与国家的关系起着调节和指导作用。马克思主义的伦理学认为经济基础决定了道德,社会发展影响着道德标准的变化,通过

伦理道德的不断发展,实现人类文明持续进步的根本目的。道德与法律一直处于相辅相成的关系之中,道德与法律通过相互补充、相互支持体现调节社会的重要作用,当人工智能介入到法律运行系统之中后,道德自然对智能司法产生影响。

现如今,人工智能广泛地应用于生活的各个领域,在智慧法治尤其是智能司法领域已经如火如荼,取得了突出成效,但也产生了一些伦理问题。首先,司法以公正为目的,司法本身就蕴含着伦理问题,这正体现了道德与法律的关系。智能司法所依靠的大数据技术带来的算法歧视、隐私权、被遗忘权等问题不仅是技术伦理问题,如果产生不公后果,则诞生法律伦理问题。其次,智能技术或机器可以用来行善,也可以用来作恶,智能既能实现司法公正,又存在损害司法公正的风险。通常的伦理学研究的是人行为的善与恶,当人工智能可以代替人展开行为时,人工智能行为的善与恶也应当被伦理学所关注。特别是人工智能占据科技的前沿高端,但其发展前景却有不确定性,人工智能所制造的社会事实并不一定符合人类的伦理要求,甚至会引起事实与价值之间的巨大裂

基金项目:本文系2020年国家社会科学基金项目“贵州少数民族聚居地方习惯法司法适用研究”(20XMX032)的阶段性研究成果。

作者简介:罗洪洋(1965-),男,贵州石阡人,法学博士,中南财经政法大学法学院法理学教授、博士生导师,研究方向为民法、法人类学、法社会学、法文化;李相龙(1993-),男,河南郑州人,中南财经政法大学法学院博士研究生,研究方向为法学理论。

缝,对此可能出现的风险问题,应当未雨绸缪,提出前瞻性的预案。

### 一、智能司法之伦理前提: 智慧知识与伦理道德的关系

人工智能运用于司法会产生伦理问题,这是人的智慧知识与伦理道德的特殊关系所决定的。人类作为万物的灵长区别于动物的显著特征就是,人有智慧知识和伦理道德而动物却没有,那么智慧与伦理到底有什么关系,一直是学术研究的重点但也是众说纷纭的。苏格拉底提出的“德性就是知识”,还认为“无人有意作恶”。<sup>[2]P65</sup>亚里士多德对此的解释是“一个人如果知道那个行为是恶的,就不会去做,没有人会明知而去做与善相反的事,除非不知。”<sup>[3]P194</sup>苏格拉底只看到知识与德性相一致的一面,却没有看到知识与德性不相一致的一面。“无人有意作恶”更是片面的,只有像苏格拉底这样极其高尚的道德君子才能够做到不有意作恶,普通的人和邪恶的人作恶多是有意。所以,一个人作恶犯罪与其知识没有多大的关联,这一点司法人员都知法,但其犯罪率竟然高于社会平均犯罪率,这也不能不引起人们对知识与德性、法律知识与守法德行的思考。知识求真,德性求善,真与善都是人类所追求的,但二者并非总是一致的。苏格拉底企图实现真善一体的思想是失败的,他自己是这种思想的践行者,他认为雅典公民大会对其判处死刑的审判是不公正的,他有多种逃生的机会,但是他都放弃了,他知道接受法律的惩罚是正义的,逃脱法律的惩罚是不正义的,他宁可接受不正义的惩罚,也不愿意做逃脱法律惩罚的不正义的事,所以他不可能逃脱法律惩罚,否则就是有意作恶,这就违背其“无人有意作恶”的信条,苏格拉底只有从容赴死这一种选择来践行他的“德性就是知识”“无人有意作恶”德性观。

正是知识与德性既有一致的一面又有不一致的一面,所以,既要实现善充满信心并利用知识实现善,又要对知识可能损害善甚至做出恶保持必要的戒备和及时拿出应对之策。智慧知识具有双刃性,既可以实现善,又可以制造恶。特别是随着科学技术的高速发展,在牛顿三定律的科学发现和电技术发明之后,科学制造的社会事实充分体现出知识与善恶之间的张力关系。就善而言,造福人类的科学技术,农

业科技使人们免于饥饿,生命科技使人们免于病亡,生态科技使环境免于荒芜,法律社会科学使人际关系和谐、公民受到善待、人有尊严的活着;就恶而言,如果科技被利益牵制和不当操纵,也可以作恶,如农药威胁食品安全,化工威胁环境安全,核技术威胁人类安全,而基因婴儿编程和克隆人则威胁人类的基因安全。因此,人类的科技必须具有去恶从善的伦理规范,诚如亚里士多德所言“每种技艺与研究,同样地,人的每种实践与选择,都以某种善为目的。”<sup>[3]P1</sup>亚里士多德还指出,每种技艺所追求的善也是不同的“医术的目的是健康,造船术的目的是船舶,战术的目的是取胜,理财术的目的是财富。”<sup>[3]P4</sup>亚里士多德虽然没有说司法所追求的善是什么,但是从亚里士多德在《政治学》和《尼各马可伦理学》对公正的高度关注,司法的善就是公正,亚里士多德把公正分为分配公正和矫正公正,司法的公正主要是后一种。司法公正对法官的智慧知识和善良正派都提出了要求,所谓知者不惑,法官即使有实现司法公正的良好愿望,但如果缺乏足够的法律知识和司法智慧,也会陷入迷雾中被各种假象迷惑了双眼,因此,司法智慧对司法公正的重要作用无论怎么强调都不为过。

智慧知识与伦理道德既有紧张关系,又有紧密关系,因此,讲德性并不能局限于道德的善恶,还要联系智慧知识的聪明与愚蠢、真实与虚假。例如,亚里士多德就认为,德性分为理智德性和道德德性,分别是教导生成的和习惯养成的,理智德性又可以分为理论理性的和实践理性的,智慧和明智分别属于理论理性的和实践理性的德性,“显然离开明智就没有严格意义的善,离开了道德德性也不可能有明智”。<sup>[3]P190</sup>即不明智就没有做善事的实践能力,当然就不会产生严格意义的善;明智属于实践理性的,亚里士多德虽然在这里没有提出智慧,但智慧既然是属于理论理性的,智慧与善的关系和明智与善的关系也是相近的,即一个人的行为没有智慧也不会有严格的善。亚里士多德不仅指出智慧与善德的紧密关系,而且也认为二者的分工是不同的,“与没有德性<sup>①</sup>的情形一样,离开了明智我们的选择就不会正确。因为德性使我们确定目的,明智使我们选择实现目的的正确手段”。<sup>[3]P190</sup>因此,现代德性的法官,不仅要具有善良的道德德性,还需具有智慧的、明智的理智德性才能实现司法公正,而人工智能司法则是法官智慧知识的延

伸而已,因此人工智能使用于司法,必然会涉及到突出的伦理问题。人类社会越发展,智慧知识越有力量,其与伦理道德的紧张关系也更加紧张,知识越有力量越能够作更大的恶;与伦理的紧密关系也更加紧密,知识越有力量越能够行更大的善。正是由于科学技术的大发展,事实与价值关系的紧密性和紧张性也更加突出,对科技的伦理审查也显得尤为必要,一些伦理审查直接转为法律规定。

## 二、智能司法与司法伦理的紧密关系:实现司法公正的伦理价值目标

从古到今,司法公正的一个很重要的因素是证据的获取,证据的获取与知识智慧有很大的关联性。古时候科学技术水平低,科学对证据的获取帮助不大,证据获取的一个重要源头是法官的聪明才智。但是法官的聪明才智获取证据至少有三方面的不足:一是不够准确。即证据可能是相当程度的真实,但根本达不到百分之百的真实,法官主要采用经验归纳法来获取证据,没有现代技术对证据呈现的准确性;二是没有重复性。法官的这种聪明才智一旦被当事人识破,就没有什么价值了,同类性质的案子总是很多,法官的聪明才智只是在最先的案子里才能够获取证据,随后这聪明才智对于获取证据逐渐失效;三是速度慢和不确定性。一个证据的确证需要另一个证据来证明,而这另一个证据又可能需要第三个证据证明,环环相扣,任何一个环节接出了问题就导致证据链的断裂甚至最后不了了之。

以古代以色列所罗门判案为例。两个妇女为争是一个小孩的母亲而诉讼到所罗门那里,所罗门充分利用了“母亲不会伤害自己的小孩”这个人之常情的经验,故意提出将孩子一分为二的平分方案,一个妇女坚决反对,另一个妇女赞同,所罗门将小孩判给反对方案的妇女。他正是从两个妇女对方案的态度获取小孩归属的证据,但是这种证据的获得具有上述所说的三方面的不足。以不够准确为例来分析,亲生母亲在通常情况下为了孩子而选择败诉,但不排除例外的情况出现,而非亲生妇女为了胜诉通常不考虑小孩的安危但也有例外,如果这两个例外发生一个或两个,那就出现两妇女都不同意方案、两妇女都同意方案、小孩母亲同意方案而非小孩母亲不同意方案这三种例外情况,那么前两种例外情形所罗门无法决

断,后一种例外情形所罗门会作出错误的决断。当然这三种例外情形之所以称作例外,就是因为可能性极小,但肯定有这种可能性,所罗门所设想的第四种情形,即小孩母亲不同意方案而非小孩母亲同意方案只是通常情形,概率再高也只是或然性而非必然性。这就表明,所罗门利用自己的聪明才智来获取孩子归属的证据,只是当时科技不发达没有基因鉴定技术所采取的退而求其次的无奈之举,这样的证据在今天看来是有很大的瑕疵的:第一,不够准确。没有基因鉴定技术的确定,最终的结果都是所罗门的推测而已,还是存在判断错误的可能;第二,不可重复性。因为这样的聪明才智运用之后就广为人知,当事人很容易找到应对之策,如果把所罗门的方案运用于今天,非小孩母亲甚至会抢着说不同意该方案以表明自己正是小孩的母亲,所罗门提出的这个方案在当时是有效的聪明之举,在今天则是无效的荒谬之举;第三,慢速性和不确定性。虽然所罗门对这个案子判断很快且具有确定性,但这只是通常情况,有一定的运气,如果出现前述的三种例外情况,要么无法作出判断,这势必要寻找新的方法断案,需要多长时间才能找到有效的新方法是未定的;或者说在所罗门当时的情况下,如果他的这个方案行不通,他也很难找到其他有效的方案。要么在双方当事人都作出例外的反应下,所罗门也因为还按照通常的出牌规则而被误导,就会作出错误的判断,从而使证据的判断具有不确定性。

但在现代社会,人工智能应用于智能司法中,对证据的获取远超人工,极大程度上保证了司法的公正。本文对人工智能司法作广义的理解,类案智能推送系统、司法大数据系统、诉讼服务系统、司法人工智能系统、法律机器人等都是本文所指的人工智能司法。人工智能在当前仍然是人的智力延伸,尽管在创造力上远不及人类,但其速算和信息处理等远远超过人类,智能机器的计算能力数亿倍甚至数十亿倍于普通人。人工智能使用于司法,在证据获取等方面具有相对于人工的以下优势:

### (一) 准确性:提升证据认定的科学性、可靠性

准确是科学的重要标准,在证据科学中,准确也是证据合法性、可靠性、可采纳性的首要标准,不准确的证据就存在着瑕疵风险。人脸识别、摄像头、大数据等广泛运用,极大程度上提高了信息的准确性,增加了能够还原案件事实真相的可能。人脸识别技术

布置于人员密集区域,能自动获取动态的人脸图像以便与公安机关的数据库进行比对,进行身份筛查。“已在公安机关的侦查活动中发挥重要作用,使演唱会上抓获逃犯等新闻频登头条,成为公安机关追捕逃犯的得力助手”。<sup>[4]</sup>对事关司法公正的证据认定和其他信息数据甄别,人工智能司法的技术和机器最具有准确性,最能够震慑犯罪分子,打消其侥幸心理并及早悬崖勒马,最能给无辜者以清白,也最能给见义勇为者以正能量的激励,从而避免做好事被讹诈的后顾之忧,这些都是容易出错的人工无法做到的。一个法官办案,对证据的认定方向或多或少地受前见、先见、成见甚至偏见的影响,这是很难克服的,例如以貌取人,凭当事人的神色面相断定可能作案或不可能作案,这就决定了证据认定的错误方向。例如,《吕氏春秋》的寓言“疑邻窃斧”讲,一个人丢了一把斧子后怀疑是邻家小孩所偷,发现邻家的小孩走路、脸色、神色、言行无不像偷斧的人。后来,丢斧的人在山里水沟掘出那把斧,再看邻家的小孩无论走路、脸色、神色、言行都不像偷斧的人,同样的邻居小孩,为什么丢斧人的前后判断不一样,就是因为前有偏见后无偏见。还有,人的感官判断不可能那么精细,所见、所闻、所思都有可能出偏,何况任何个人都不能全天候不间断地工作,这也会在需要持续性的工作上留下盲区,既使有人员换岗,但换岗这个交接点也是容易出问题的,从而对证据的认定失准,影响司法公正。人工智能技术和器械就可以克服自然人在这方面的缺陷,可以持续工作且无失误,协助司法机关准确获取证据。“一方面,人工智能通过自动识别提取法律事实和情节、自动推送关联法条、争议点自动整理以及裁判文书自动生成等来辅助法官进行部分司法决策。当法官部分司法决策被人工智能辅助时,直接压缩了法官直觉判断运用的空间,从而相应地也减少了产生偏见的可能性。另一方面,人工智能还可以通过偏离预警和裁判预测等司法辅助系统为法官决策提供便捷有效的检测工具,从而能够有效地防范和控制法官自由裁判的风险”。<sup>[5]</sup>正是这种准确性确保案件事实清楚和证据确凿,确保司法公正。

(二) 重复性: 智能司法保障法律事实认定的稳定性、普遍性、重合性

人工智能技术或机器都是通过科学的定量定性分析生成的,其使用具有普遍性,不因人因时因地而

变化,可以在任何人、任何地点、任何时间重复性使用。可重复性是一项自然科学的科学性评价的核心标准,它基于自然科学的不因时因地因人而异,排除了主观因素的实验、检验、化验所发的因果事实,具有普遍性和客观性。只有重复性才能确保一项科学发明或发现不是出于孤立的巧合,更不是源自数据上的造假。而是具有客观的、普遍的规律性,只有客观、普遍、规律才与可重合性是融合的。重复性使用是科学技术的重要特征,没有使用的重复性就不配称科学技术,而且成本很高难以为继,因为不可能对每一个新人新案都去采用新的技术或对策。就像上述所分析的所罗门的办案方法多少有一点心理战的意味,绝不是科学的方法,这个方法的套路一旦被识破就无效了,所以不能重复使用。一个法官再聪明,他针对个案采用的方法再高明,终究是权宜之计,因为一旦被对手识破就很容易被反制。人工智能技术或机器具有刚性和稳定性,对侦破对象是可以重复使用而不被反制,除非对手采取一些破坏性的措施,如破坏摄像头、伪装欺骗人脸识别、黑客攻击大数据、篡改机器人的程序等,但这些所谓的反制措施力度极其有限,这些破坏性措施对人工智能技术或机器使用的重复性并不能构成实质性威胁。因此,尽管人工智能技术和机器也在不断开发升级和更新换代,但在升级换代的周期内,是可以有保障地重复使用的,这就大大节约了人工智能技术或机器的使用成本,提高使用效率,拓宽司法资源,提供了实现司法公正的资源基础。

(三) 快速性: 智能司法加速案件审理速率和节约司法成本

一个证据的信息元会多至数以亿计或更多,而人工处理信息的速度慢且不精准,一旦出错就要推倒重来,更加减缓信息处理速度。所以,对于信息元海量的案件,靠单纯的人工操作来准确认定证据犹如大海捞针一样概率很小,运气的成分很大。白银连环杀人案<sup>②</sup>之所以二十多年没破,就是因为人工能够处理的信息量相对于破案所需要处理的信息量是很少的,只能寄希望万里挑一那样的比对运气,这样的运气一直没来,所以案件无法侦破。而科学认定证据是讲必然性的,机器的计算能力是人工的数亿倍甚至更多,大数据处理信息速度是人的数十亿倍甚至无法计算的倍数,因此,数据库对于案件所需要处理的所有信息元都能够快速且准确地予以比对,因为人工智能的快

速能够实现对所有信息元全覆盖处理,进行现场证据与资料库数据的比对。<sup>[6]</sup>就白银连环杀人案而言,如果没有数据库比对,单靠人工查找出和他远房堂叔的DNA,全国这么多人,逐一对比哪怕除掉女性和小孩,具有作案嫌疑的可能信息元也有数千万之多,人工哪怕是数十人比对,一天又能处理多少信息元,故这个案件拖延这么久才告破。可以说,假如没有数据库的投入使用,这个案件的侦破依然是遥遥无期的。或者说,假如他的堂叔不犯案,其DNA没进入数据库,这个案子也是难以侦破的。这样的案例决非个案,而是具有普遍性,因为人工智能的这些技术和器具,把社会变成了透明的社会,如果说在此之前犯罪分子还可以伪装得像好人一样,现在则是彻底现形,这样整个案子一下子就事实清楚、证据确凿,实现个案正义就有了最为根本的保障。于是便有了,北大弑母案的吴谢宇虽然人间蒸发似的消失3年之久,还是被机场的人脸识别系统捕捉到行踪而落网;多起命案在身的劳荣枝虽然已经逃亡20年之久却还是在大数据分析下显现了原形。人工智能技术和机器一旦投入使用,那些极其狡猾在常规状态下能够伪装和躲藏的犯罪嫌疑人的好日子随即到头了。而全社会所有的人一旦对人工智能有了正确的了解,意识到自己在人工智能面前是透明的而不再有伪装的余地和藏身之处了,一旦犯案必被抓捕,那么还犯案就不仅仅是可耻的,更是愚蠢的,明智的选择就是悬崖勒马,远离作奸犯科之事,社会风气很自然地就会好起来。

各地智慧法院建设中推出的庭审智能语音识别、量刑辅助、法律问答机器人都有助于快速甄别和处理信息。例如,在讯问过程中,犯罪嫌疑人被问及是否认识被害人,作出否定的回答时,侦查人员需要花费较长的时间去核对犯罪嫌疑人所言是否属实。而采用人工智能技术的电脑屏幕可以当场显示出犯罪嫌疑人与被害人的既往通讯联系,犯罪嫌疑人立即被揭示出说谎,这就加深了其犯罪的嫌疑。人工智能参与司法尽管在目前仍然是辅助性质的,可信息处理的速度和准确度相对于人工无与伦比的优势,大大提高了案件信息的处理速度和准确度,对于尽早查处案件真相和及早结案,防止正义迟到,维护司法公正具有特别的意义。

(四) 抉择性: 智能司法为司法判决提供多元、可评估依据

司法判决是确定性与正确性的统一,那么人工智能在司法中能否作出正确抉择,这既是设计者的目标,也是设计者的疑问。正是这对智能机器更高的要求,尤其是机器人学习能力,也许将来的学习社会与现在的学习社会不同之处包含学习的主体不同,现在的学习主体只能是人,将来的学习主体有可能包含机器人,至少机器人的设计者在向这方面努力。机器学习是“让计算机修改或适应它们的行为(无论这些行为是作出决策,还是控制一个机器人),这样,它们的行为会越来越精确,而行为的精确性是这样来测定的,既定行为能否作出正确的反映”<sup>[7]P4</sup>。对于司法判决的裁判,我们现有两大法系——判例与成文法系都对裁判的正确性有着不同的观点。我国主要实行成文法,但是两大法系在我国逐渐展现出相互融合的趋势,例如我国的司法改革进程中不断加深判例对司法裁判的参考作用,最高人民法院还专门汇编司法指导判例。随着社会的发展,其复杂程度同样加深,法律规范会增加,同样判例也会数不胜数,那么未来人工智能势必成为司法裁判的重要助手。在司法判决中,大量的判例与法律规范势必会出现价值相竞、内容相矛盾等问题,在巨大的司法工作量面前人工的裁判或许会出错,人工智能程序可以帮助司法裁判分析案情。在人工智能还不具备价值判断的当下,面对复杂案件还需要法官作为主导,但当人工智能具备了学习能力、应变能力、判断能力之后,可以显著提升司法裁判的效率以及抉择性。

还要强调一点的是,人工智能在实现司法公正中发挥着越来越重要的作用,但人工智能至少目前还是在人的控制之下的,人工智能的模拟智慧越发达,其使用者人的智慧也是越发达的。尽管像所罗门单纯依靠人的智慧破案的日子已经一去不复返了,但这绝不是降低司法工作人员的智慧在破案中的巨大作用。恰恰相反,人的智慧和人工智能的智慧是相得益彰、比翼双飞的,人的智慧或许是无形的,却是起点、是方向、是筛查过滤的第一关,人的智慧充分发挥了,就能避免很多毫无方向感的试错所带来的司法资源消耗,大大提高破案速度。例如,一个案件锁定了犯罪嫌疑人及其作案的车辆,但犯罪嫌疑人反侦察能力很强,已经清理销毁了可能作为犯罪的证据,案件进程一时处于停滞状态。一个善于溯因推理的法医看到车内的玻璃上有一只死蚊子,他的推理过程如下:玻璃是

竖放的,上面的蚊子一定是粘上去的?粘的东西正是蚊子的血?粘蚊子的血来自于蚊子被拍死?蚊子会飞怎么被拍死,可能是喝血太多飞不动了被拍死的?蚊子的身体里有犯罪嫌疑人的血?把蚊子的血检验并比对犯罪嫌疑人的血检验,就可能成为锁定嫌疑人的证据之一。所以,司法有了人工智能辅助,不仅没有降低对司法工作人员的智慧要求,反而是提高了这一要求,不仅对法条的理解不能够机械司法,对证据的获取也同样不能机械司法,司法提出了对公正伦理和智慧技艺的双重要求。

(五) 无私性: 智能司法更加具有自主性、中立性

尽管我国有相关法律对法官、检察官、律师等法律人的职业伦理作出了严格规定,把司法公正作为明确的目标,把实现个案正义作为评价指标,要求法官公正无私,对徇私枉法的法官严惩不贷,但并不能说明每一个案件都能保证公正性。这其中的一个重要原因是司法工作人员的私心作怪,经不住利益的诱惑,利令智昏,知法犯法,枉法裁判。尽管有《法官法》这样的法律明文规定,司法队伍犯案的并非少数甚至包括司法高官,原因在于,司法人员也像正常人那样需要利益维持生活,如果不能够严格自律和遵守法纪,手中的权力就容易被盯上,成为行贿和收买的对象,走向以权谋利的犯罪道路。但是智能机器人仍然是机械的,是无机体,不是像血肉之躯的有机体,既没有情感也没有利益需求,机器人是没有人情世故和偏见的,也不可能被收买,所以能够做到秉公办案、不为私利、不徇私情;当然也没有人会愚蠢到向机器人行贿的地步,因此,人工智能司法不可能徇私枉法,具有实现司法公正的内在必然性。

### 三、智能司法与司法伦理的紧张关系: 损害司法公正

智能司法与司法伦理的紧张关系比较复杂,主要表现在如下几方面:

#### (一) 算法歧视对司法伦理的冲击

在人工智能司法中,算法决策在刑事司法中应用较广,就是基于数据库的个人数据利用算法对其犯罪或重新犯罪的可能性进行评估。人工智能的算法决策的目标是克服法官个人的主观性所带的偏见和歧视,但是算法决策本身也会产生算法歧视,而且是一种具有相当隐蔽性的歧视。人工智能和互联网的深

度融合,数据成为具有全新意义的新概念,数字社会、数字经济逐渐获得社会的认可,在人工智能和互联网平台上,强者恒强、赢者通吃、弱者恒弱、输者通赔成为带有歧视性的算法的必然产品,人与人之间的鸿沟体现在他们在数据上的鸿沟。尽管数据是个人产生的,互联网平台却可以利用个人产生的数据进行商业活动,因此,数据的所有权并不再属于产生它的个人。恰恰相反,互联网平台上的数据受制于算法决策后,尽管仍有客观性、平等性的外表,但却可能既不客观又不平等,原因是个人在互联网面前透明化,已经没有什么隐私可言,存在着被算法决策朝着各个方向利用的可能性。有的方向对于个人是有利的,有的方向对于个人是不利的,这不利的方向就形成歧视。本来是用来审理案情的数据,一旦在互联网上,其使用的方向就难以控制了,甚至个人信息的泄露被用来人身攻击、人身控制、窃取商业秘密、盗取诈骗个人钱财等违法犯罪活动。

算法歧视不一定是设计者故意创造的,而是这些数据指向了某种歧视,并导致这种歧视的自我编程和蔓延,反而给人一种没有歧视的感觉。比如在美国通过 Google 搜索非洲裔人,会显示许多负面信息,虽然是客观事实的呈现,但无疑是对搜索者产生警惕非洲裔人的影响。当你面对非洲裔人的求职者或申请者时,出于对安全的考虑会想调查他的犯罪背景,当这种偏见产生,会导致非洲裔人生存难度增加而被迫犯罪的恶性循环。<sup>[8]</sup>这种偏见导致雇主不愿意雇佣非洲裔人,于是非洲裔人更加的贫穷,部分人为了生计真的走向犯罪,这反而印证了雇主对非洲裔的判断,甚至认为自己对非洲裔人是洞察而非偏见,非洲裔人有了这样的不良印象后会在求职中处于不利地位。数据本身只是客观事实,人们对数据的解读一定是从价值的视角出发,与数据的产生者的交往带有价值倾向的选择,这种选择性就是变相的歧视。

司法审判中,如果人工智能所呈现客观的双方当事人数据后,该数据是为协助法官进行司法活动,实现司法判决的公正,可因为数据的歧视性影响了法官主观的判断,会产生对裁判公正性不利的影响因素,同时在数据逻辑的影响之下,法官的自由裁量权会受到制约。例如刑满释放人员的数据进入系统,那侦查人员在甄别犯罪嫌疑人时,刑满释放人员就会被当作有前科而成为重点关照对象,法官对有前科人员的审

理也是格外细心,且在量刑时以累犯加重处罚。虽然有前科的人员再犯的可能性比无前科的人员首犯可能性要大,被当作嫌疑对象的重点合情合理,但被不合理怀疑和监视,放松对其他无前科人员在同样情节下的怀疑和监视,就具有歧视性。对累犯的加重处罚不仅是法律的规定而且也具有合理性,但法官在审理累犯时如果带着仇视、偏见、鄙夷等先入之见,被数据库里的累犯记录所刺激,那加重的处罚可能就不仅仅是累犯的身份,而是该累犯的案情细节都有可能被放大了,对累犯的歧视已不经意产生,尽管这与法律对累犯的加重处罚是完全不同的两码事。虽然侦查人员和法官都不是有意为之的,而是前见、偏见和歧视所带来的惯性,但这种惯性不仅具有伪装性甚至使人们误以为或不承认它是歧视,而且这种歧视意味着区别对待,与司法公正所要求的平等对待严重对立。算法歧视不仅冲击司法伦理,而且具有伪装欺骗性,更应该引起重视。

## (二) 改变司法公正的性质和法官的主体性地位

司法公正体现在个案正义的实现上,法官是办案的主体,法院的案件不管多少都是一个一个案子办下来的,这与工厂的产品是流水线一批一批生产出来的是明显不同的,而人工智能的使用却把这明显不同变得模糊。当法院审判案件如同机器工作一般,案件被输送进法院,通过程序进行分解、分类、计算、审判,将会削弱审级制度,统一程序下案件的审判结果呈现一致。证据获取、审查、计算、量刑、审判、执行的一体化会让整个司法系统统一化。这些统一会使法律的发展跟不上社会发展的脚步,对现代的法治发展产生巨大的影响,这是人工智能司法发展道路需要思考的重要环节。<sup>[9]</sup>必须承认,法院办案和工厂生产是不同的,工厂的生产强调标准化,可以用流水线生产出批量的相同产品。法院的每个案子都存在个性的不同,因此案件不能像产品那样标准化,也无法进行流水线作业,而现在的人工智能司法确实有朝此方向发展的趋势,如果能变成现实,那就不存在个案与个案正义了,这是严重违背不同于工厂生产的司法办案规律的。机器人法官的设想也是可怕的,尽管机器人在智能的某些方面上与人有某些相似性甚至远远超过人,但从根本上讲,机器人不具有人类智力最为核心的创造性和自主性,更不用说机器人不具有人类的道德情感。机器人在司法中只起辅助作用而不是主导作用,

更不能取代人成为司法的主体,所谓的机器人法官这个概念是凭空设想和不成立的。而且,为了实现司法公正,就要坚持司法责任原则,“让审理者裁判,由裁判者负责”,法官对错案负有终身责任。如果机器人能成为法官,那它是司法主体吗?它能对错案负责任吗?回答是否定的,因为责任是与惩罚相联系的,机器人是无机体,没有痛苦和快乐、没有情感和理智、没有生命和财产,对任何惩罚都没有反应机能,也就没有办法惩罚它,也就无责任可谈,所以不能成为责任主体。正如黑克所言“现代的法官绝不是一台法律机器,而是在很大程度上充当立法者的助手,有着更高的自由,但相应地也负有较重的责任”。<sup>[10]P31</sup>如果机器人取代法官成为司法主体却不能承担责任,办了错案也无法追究责任,那就违背司法原则,对司法公正和司法伦理的冲击是巨大的。

## (三) 大数据和机器人出错所带来的伦理价值效应

智能的大数据和机器虽然不像人那样频繁出错,但也有出错的可能,有些出错并没有造成太大的危害,而有的机器出错却会造成伤害。如自动驾驶汽车,因为出错造成人员伤亡,且造成伤害难以得到及时纠正;机器人充当法官办案,出错导致办案证据被毁、办案方向背反、案件当事人误认甚至伤人等,这些都应该在我们的考虑范围之内。数据和机器人出错主要包括以下几种情况:

1. 设计出错。智能设计需要排除伦理偏见,与现实接轨。设计出错分为两种:

(1) 带有伦理偏见的设计,这就直接影响司法公正。例如,2015年谷歌公司的图片识别服务错误地将某些黑人标记为大猩猩后,被批评种族主义事件的谷歌公司立刻采取措施在搜索结果中删除了这一标签。<sup>[11]</sup>再如,域外刑事司法领域的人工智能评估系统,基于算法本身而形成了针对有色人种的偏见,误导法官、检察官和警察随之产生对有色人种的偏见。<sup>[12]</sup>今年的弗洛伊德被白人警察暴力执法致死的案件报道最初还有一种说法是,弗洛伊德因涉嫌使用假币并拒捕,所以警方为了制服弗洛伊德而意外致其死亡。这两处不同的新闻媒体视频和报道,使人们对警察与弗洛伊德的权责认定和司法公正发生不同的认识。如果智能系统有选择地对报道浓墨重彩而对视频轻描淡写,那必然对本案的信息处理缺乏客观



性,而是主观选择性和偏见。因为算法系统的高新性,普通群体很难了解,那么设计者就存在着带入偏见、武断等不公的因素,资本或政治对算法的参与更加导致算法带有一种倾向性,这种倾向性体现在算法歧视,面对这种高新技术的压迫,反抗的难度大大加深。<sup>[13]</sup> 尽管智能技术或机器本身并无情感和偏见,但智能技术和机器不是自我生成和编程的,是人设计和编程的,只要设计者自身有偏见,那么智能技术和机器在使用中不可避免地具有偏见,而偏见与司法伦理的公正要求是针锋相对的,凡是司法存在偏见的地方,司法公正就不能实现,司法伦理的底线就受到严峻挑战。

(2) 设计不合理 影响效率 挤占司法资源 间接带来司法不公正。设计者和使用者是分离的,即设计者多是计算机方面的专家,而使用者多为对计算机技术并不精湛的法律人,二者不能够很好地衔接起来。有很好的设计由于使用者专业知识的欠缺而无法使用,让设计者来使用也不现实,原因在于,设计者缺乏法律知识、法律思维和法律精神,设计者使用只能是机械司法,破坏司法公正。正如前述,再发达的智能技术和机器都不能成为司法的主体,影响司法公正的很多因素是无法设计的,所以非但机器不能,就是机器的设计者也不能取代法官成为办案主体,不能成为他所设计的机器的办案主导使用者,最多是辅助者的身份。由于个案各不相同,统一证据标准也面临着设计和操作上的困难,机器数据的标准化设计与个案数据的标准不一致具有不可调和的矛盾。一些设计本来是为了减轻工作人员的负担,但由于操作繁琐、系统僵化或其他没有多大必要甚至干扰性的流程,反而增加了司法工作人的负担。

2. 故障出错。智能系统的故障破坏力强且纠正难度大。虽然设计没问题,但是出现故障,导致机器人的认知与设计出现偏离甚至相反的认知,如设计的是举证相关的真实数据,出现故障后有可能是举证无关和虚假的证据。尽管机器人或数据系统出错的可能性较小,但出错后常常带来重大事故后果,如曾经发生的机器人杀人<sup>③</sup>,设计没有杀人的指令,一旦出现故障,机器强大的力量会误伤人们,机器人本身没有是非观念,只按照出现故障后的新指令操作。系统和机器的故障出错由于得不到及时和有效纠正,比人工出错更加可怕,那些具有杀伤力的机器人在故障出

错时常常已经脱离了人的控制或人无法对其进行控制,这是目前工业安全的一大隐患。不仅仅工业机器人具有风险,智能系统已应用于社会生活的方方面面,数据常伴现代社会,如果是智能系统的司法大数据出现故障,波及的范围更广,数据失真甚至被篡改,对司法公正的潜在风险非常之大。

黑客攻击导致数据、程序、指令被更改的故障在故障出错中占比很高,黑客通过高明的伪装将病毒植入被侵系统,这些病毒可以篡改程序为黑客所用。这样的系统被侵入后,名义上的主人未变,实际上已经变了,即系统的服务方向已经发生根本改变。如被黑客攻击的自动驾驶汽车不仅刹车失灵,方向盘也失灵,汽车则完全失控停不下来,一个刹车和方向盘失灵的自动驾驶汽车在完全失控的情况下,其破坏是毁灭性的。如果司法的办案系统或辅助的机器人也是这样的故障出错,会产生违背司法公正的后果,如果发生司法数据的泄露还会伴随侵权责任。

3. 应变错误。随机应变是人的能力,理论上,智能机器经过深度学习为应对复杂环境也应该具有这种能力,且在很多环境下,人类对人工智能也确实提出这样的要求,但智能机器要想达到这样的要求也是困难重重,程序化的智能不具备应变能力。比如,人们已经遇到这样的问题,广泛使用的智能无人机,航行和射击指令完全由人工智能掌控,当本应该是恐怖分子聚集地却出现平民等突发情况,则是对智能程序判断的更高要求,射杀恐怖分子可能造成平民伤亡或者房屋财产损失。这是因为,任何一种针对具体动态情况的设计都是粗线条的,不可能涵盖具体情况的不间断变化,机器人要根据这种变化的情境作出随机应变的对策。机器人根据指令活动,而在情境变化情况下,指令与现实相违背需要作出应变,这种应变能力所属人类而非机器。就人自身而言,应变能力除了自身的知识结构和思维能力外,还有对经验的利用,而机器人是无经验可言的,所以可能出现应变出错的问题。对人工智能司法的过于乐观也源自对司法的逻辑崇拜和经验贬低,逻辑是输入的,所以智能机器的逻辑能力很强;而经验是生成的,所以智能机器的经验意识很弱,即使智能机器的深度学习能够获得一部分经验也是局限的。在司法中,逻辑和经验都很重要,只是分工不同。逻辑代表了法律知识、法律训练和法律技巧,经验体现了实践积累、社会知识和社会



阅历。逻辑可以通过输入性的学习培养,而经验却是一种在历练中生成的技巧。<sup>[14]P35</sup>智能司法和人工司法的比较区别是:胜在逻辑上和输在经验上,这也是智能司法无法取代人工司法的原因之所在。

伽达默尔认为,法律不仅仅是将法律条文应用于相应案件之中的机械行为,这种单一的逻辑分类工作只是司法的一小部分。<sup>[15]P685</sup>美国大法官霍姆斯指出,创造法律的关键在于经验,每个时代的伦理道德、对美好生活的需求、政治环境的影响、法官或立法者的偏见对法律制定的影响都大于三段论的推理。<sup>[16]P1</sup>那种认为人工智能代替法官的想法和方向,都是低估了人类法官在司法活动中对经验智慧的运用和随机应变的能力。人工智能的智力主要体现在计算和数据处理上,来源于程序设计,而法官的智慧来自知识、经验和临场的洞察,甚至为实现正义的道德情操的激励,这些来源都是人工智能所不具备的。因此,尽管人工智能的计算和数据处理可能数十亿次倍于人工,但这是量的比较,不是质的比较,不能说明人工智能的办案质量强于法官,更不能改变法官的主导地位,不是人工智能成为办案主体的佐证。虽然程序里也有知识,但这种知识是输入的,与法官的生成知识具有不同的性质。即,他不能在设计的知识外创造新的知识,而法官的生成知识犹如肥沃的土壤,能够基于情境产生新的知识,如所罗门断案等。尽管随着高科技运用于司法,法官对经验的依赖性减弱了,但经验对于法官对证据的甄别和方向的判断上作用并没有降低。

#### (四) 人工智能本身不具备价值判断

司法判决要求事实清楚、证据确凿,但是单纯的事实并不能形成判决,案件事实只是司法判决的三段论小前提,大前提则是法律规范,而法律规范是蕴含着价值的。现代法治并不能仅仅停留在依法治理的事实层面,还特别强调所依之法是良法的价值层面,现代法治讲究良法善治,从价值判断来讲,恶法并不在法的涵摄范围之内。在司法过程中,对价值的判断指引着司法判决。如英国杜德利案、二战德国告密者案、许霆案、唐慧案等等,这些案件都存在着对司法、正义、论证的价值判断,这种价值判断受到社会道德伦理、经验理性、时事政策的影响。这些影响因素智能司法不能体现,人文关怀与伦理道德是智能司法不能计算出的。<sup>[17]</sup>司法逻辑崇拜和机械司法之所以受

到批判,就是因为价值判断的缺位从而导致司法不公,背离司法目标。拉伦茨提出“学者们共同认为,案件事实满足某法规的构成要件,使用该法规进行裁判的逻辑过程对法律的发展并非意义重大。真正使裁判公正的是法官在面对案件时,包含价值判断的考量。”<sup>[18]P33</sup>随着社会的发展和法律价值观的变化,人们对法律特性的认识也发生变化,强制性尽管仍然是法律的重要特性,正当性更是法律的特性,是现代法律最主要的特征。近年来发生了许多被大众所关注的热点案件,如于欢案、昆山龙哥案、电梯劝阻吸烟案都体现了价值判断对司法判决的影响,这些案件推动了法律保护公民自身权利的发展。<sup>[19]</sup>价值判断不仅是司法判断的方向盘,而且具有情境性的意义,即价值不是抽象的、可提前设计的,而是具体的、相对于实际情况的。由于智能技术和机器都具有普遍性、超情境意义,并不具有特殊性、情境性意义,所以,与对司法公正非常重要的价值判断近乎是绝缘的关系,这就表明,以逻辑见长的智能技术和机器,只能对证据的法律事实进行认定,却不能对证据的法律价值进行判断,而对司法公正影响最大的一些疑难案件,价值判断才是焦点之所在,智能技术和机器由于缺乏价值判断就引发其与司法伦理的紧张关系。

#### 四、人工智能司法伦理风险的应对预案

上述分析了人工智能司法与司法伦理的两种关系,对紧密关系下的司法伦理良性互动需进一步加强,而对紧张关系下的司法伦理恶性风险则需要未雨绸缪,作出预案,且后者更是这两种关系中需要重点关注的。

##### (一) 认清人工智能司法的伦理缺位,加强伦理审查

人工智能作为高新科学技术,与其他高科技一样都具有既能造福人类又能危害人类的两面性,甚至对人类的造福越大,对人类的祸害也对应性地越大,人类对此应当保持必要的警觉。为了使科学技术的发展符合人类伦理道德的要求,社会各界纷纷组建了伦理委员会,如我国组建的国家科技伦理委员会、微软等大型科技公司设立的伦理审查委员会。现如今,机器人已经可以替代人类的许多机械化工作,世界经济论坛报告显示,2025年智能机器将替代8000多万个工作岗位,同时又会创造9700万个工作岗位。智能机器所替代的是日常生产生活的工作,具有创造性的工

作人工智能只是作为辅助,为人类提供数据上的支持。人工智能算法决策只是阶段性和暂时性的,是为人类的决定提供信息数据上的参考。<sup>[18]P8</sup> 尽管科技本身只是知识性的,与伦理没有关系,但科技的破坏性足以引起巨大的伦理问题,在科技大发展的今天,科技伦理的发展也是方兴未艾。很多科学,在知识上已经不是问题,在伦理上却是大问题,如克隆人和婴儿基因编程在科学上已经通过了,却无法通过伦理审查。我们在为人工智能司法所带来的正面性而欢欣鼓舞时要一分为二地看问题,对人工智能所带来的负面性而保持冷静的头脑和忧患意识。当今很多科技伦理学者在人工智能的阴暗面上达成共识,这里的阴暗面就是指人工智能的伦理问题而非智识问题,因为人工智能的智识已经非常发达充足了,这肯定是明亮面。鲜明对比的是,人工智能的伦理近似一片空白,伦理才是人工智能的阴暗面,司法对伦理提出了很高的要求,这就导致已经很发达的人工智能使用于司法时却不可避免地出现伦理盲区。目前的机器人只是将智识规则编程进机器人,机器人只是遵循逻辑准则,这是远远不够的;人工智能还要被编程进伦理规则,机器人必须遵循司法规范。机器人的伦理审查,就是将不符合人类价值的机器人活动强行禁止,避免其对人类个体和社会产生破坏,并以之为基础,最大程度地增加其对人类的包括伦理在内的各方面助益。

伦理审查体现在程序设计上,如果程序设计上就颠倒黑白,以善为恶或以恶为善,或者程序设计的活动与人类伦理发生严重冲突,就无法通过伦理审查。例如,故意杀人和伤害构成犯罪,是法律所需要打击的,这是惩恶,而正当防卫所造成的故意杀人和伤害并不构成犯罪是法律所要保护的,这是扬善,而无论惩恶抑或扬善,既是一种事实,又是事实所蕴含的价值判断,法律编程如果放纵甚至保护故意杀人伤害的,就是以恶为善,就无法通过伦理审查;法律编程如果无视甚至敌视正当防卫,就是以善为恶,也无法通过伦理审查。

## (二) 坚固法官主导地位

必须承认,智慧只是人的一个重要特征,情感和道德是人与动物、其他机械区别的另一个特征,这是人工智能所不具有的,即使智能机器人也不具有。何况,人工智能只是在记忆、速算、信息处理等方面远远超过人,但这些固然是智力的组成部分,却远远不是

最重要的组成部分,智力的最重要特征是创造力,即创造出比自己原先更先进的东西,而再先进的机器也是没有创造力的,只能制造出与自己同等智力程度的东西,所以,人工智能只是一个名词概念,不能够进行望文生义的理解。道德并不来自理性而是情感和直觉,机器人作为器械是无机体,既无情感又无直觉,无法产生伦理道德,最多是人操作编入不得为或当为伦理程序,即将各种法律事实的价值代码化以植入人工智能系统。但是需要避免设计人员将个人的主观偏见、情感好恶、身份歧视等带入人工智能系统,以确保系统在运行时能够作出一定程度的符合司法伦理的决断。价值代码的编程远比智识代码的编程复杂,不确定性因素更多,很难保证决断完全符合司法伦理,严格说,人工智能系统即使输入价值代码,其价值判断也是低层次的和不可靠的。

我国的司法体系之中非诉讼法律制度是智能司法不能解决的问题之一,调解是需要依靠法律规则、公序良俗、伦理道德相互结合,使轻微案件通过协议得以解决的方式,主要体现在现代社会的能动司法之中。司法的能动性体现还需要依托于法官来实现,能动司法与智能司法之间的关系需要厘清。在社会转型时期,传统司法的被动性规律被质疑是否完全符合经济发展的需要,结合我国现实国情与本土司法属性,人民法院要发挥司法的能动性,正确适用自由裁量权,灵活司法、能动干预,做到服判息诉、案结事了,积极的发挥治理作用。<sup>[20]</sup>在发挥司法的能动性同时,不能违反司法公平公正的基本原则,能动司法的本质是扩大司法的管辖范围,充分地发挥法院处理纠纷的作用,有助于全面推进依法治国。但司法判断应该是独立的、中立的、客观的、被动的,能动司法的主动性是否会破坏被动司法的理性判断原则值得我们考量。智能司法的优点恰恰可以弥补能动司法主动性所带来的司法不公的隐患,人工智能在数据的采集、分析的客观属性,使法官可以清晰地对证据进行审查,并且在大数据的帮助下,同案不同判的问题迎刃而解。同时能动司法会加大司法成本,如果推动司法的能动性,势必需要智能司法的辅助从而节约成本。

法律需要贴合人们的生活,是为人们追求美好生活所服务,只有人通过实践产生经验作用于司法之中,这样的司法才符合人们生活的需要。同样,司法活动所产生的经验,是为下一次更好的司法活动提供

支持,这样的司法经验需要法官通过判案产生,经验司法是智能司法所不能涵盖的。比如马锡五审判方式<sup>④</sup>,司法的根本目的是审判让人们所信服,司法所体现的价值被人们所接纳,人工智能所蕴含的价值判断只能是设计者的程序输入,所以,智能司法不能代替经验司法。由于社会处于一直发展的状态,没有任何一部法律是完美的,总会产生现有法律规范所不能涵盖的行为,这就需要人们通过对以往案件经验的梳理,对新事物、新行为进行解释。不管是现代司法广泛的管辖范围需要法官通过经验司法调解,还是未来社会不断发展所产生新的法律规范时需要司法经验的指导,都离不开法官的经验。智能司法可以通过数据的记录、分析,帮助法官学习、积累经验,从而实现更好的司法活动。

法官有职业道德要求,法官实行错案追究制,而机器人没有意志,所以没有权利、行为、责任等能力,也就无所谓职业道德要求,办了错案也不能被追究责任。当然,目前机器人办案仍然是辅助的性质,法官仍然是办案的唯一主体。如果机器人有一天能够发展到有意志和权利、行为、责任等能力时,那就具有职业道德要求,办了错案也能被追究责任,但这个假言命题基本上不会成真的,因为如果机器人有一天能够发展到有意志和权利能力、行为能力、责任能力,那他就与人无异且在抵抗和生存能力上远胜于人类,人类就有可能被机器人主宰甚至毁灭,也就没有必要去讨论机器人办案的主体资格问题,那早就成为不是问题的问题了。因此,实现司法公正,再发达的人工智能也只是起辅助作用,法官才是决定性的和主导性的,法官的职业道德、法律精神、法律思维是实现司法公正的根本保证,提高法官的素质尤为重要。

### (三) 规制算法歧视

算法决策有公平和歧视两个方向,前者促进司法公正,后者阻碍司法公正,因此要对算法歧视进行规制,清除阻碍司法公正的算法决策。例如,算法歧视违反平等原则,侵犯公民个人信息权和隐私权,违背法律面前人人平等的基本原则。如果司法大数据对穷人、弱势群体、年龄、怀孕、性别、民族、信仰或前科等身份具有歧视性的影响,应该进行调查和审查,确保每一个公民的司法数据在法律面前都受到平等对待,确保每个公民的平等权。算法歧视具有隐蔽性,社会生活中算法歧视不易让人发觉,人工智能时代人

的身份信息通过泄露被记录、测算、分类,这种算法带有明显的歧视性而不被人发觉。通过这种带有歧视性的算法分类会对某些特征人群、地区人群产生不公正待遇,还会通过算法歧视进行诱导行为,表现出对人的侵权性。算法歧视的隐蔽性致使人对与被歧视的事实认识不清,是对知情权的一种侵犯。算法歧视规制的第一步应该是保障个人隐私权、被遗忘权的前提下,当事人知情且同意后再进行大数据的采集。不得在不知情的情况下采集、运用个人的数据信息。

算法技术在被应用于智能司法前应被审查与评估,算法可谓是人工智能技术的核心,如果需要人工智能介入司法体系之中,离不开对算法技术的运用。由于算法技术以信息数据为基础,信息数据来源于社会公众,当被司法体系用于智能司法,需要司法机关与社会共同进行审查与评估。司法机关对算法技术的审查应从合法性、公正性进行,算法技术的合法性是对司法程序正当性的保证;智能司法的根本目的是实现司法判决的公正,故,算法技术应当保证对司法判决产生的影响符合公正性。社会对算法的评估或被其高新性所限制,可以通过专业机构、学者专家、相关用户多方参与,对算法技术的专业性、适用性等进行评估。只有算法技术通过司法机关的审查,社会多方主体合作评估结果符合公众利益与价值观,才可以被应用于智能司法。

司法诉讼是对算法歧视的重要规制。受歧视者可向法院提起诉讼,要求改变歧视性的算法决策和给予赔偿,对于这样的诉讼纠纷,法院应从维护司法公正的角度支持原告的诉讼请求,作出公平判决。还要加强算法的伦理审查,对算法的公平性、平等性检测应在设计阶段就介入和审查,只有通过审查的算法才能进入数据库使用。国家还可以制定《怀孕歧视法》和《就业年龄歧视法》等规制算法歧视。

除了对算法的评估与法律规制外,算法应当被解释。公民需要知道适用法律的依据,而不是简简单单一个算法的结果。现在的算法存在技术黑箱,当事人不知道算法决策的依据、机理,如果这样的算法技术应用于智能司法,那么是对公民知情权的无视。<sup>[21]</sup>算法应该被解释,当事人需要了解参与自身案件的算法技术,有权对算法不合理的地方提出质疑。如果算法处于一种神秘的地位,那么就是与当事人相隔绝的一种技术,这种技术的正当性无人论证,同时会被当作

影响司法公正的工具,所以算法应当被解释,应当被知情,允许被质疑。但考虑到算法是一门高深的技术,算法的设计者也对算法运作过程无法解释,就算能对算法解释也并非所有人都能理解,那么算法在司法判决中的适用需要得到当事人的同意。

#### (四) 确保当事人的人格权

智能司法在进行辅助司法或参与审判时都面对着当事人的人格权,在没有智能司法的年代,一些冤假错案致使当事人丧失数十年的人身自由,甚至有些冤假错案剥夺了当事人的生命权,这些冤假错案严重损害了当事人的人格权。我们对法律的追求正是为了避免在将来再出现冤案错案,而智能司法产生的根本目的应当是保证司法公正,不侵犯当事人的人格权。首先,智能司法只有出于保证司法公正的目的才能避免出现侵犯当事人的人格权的审判,但智能司法在数据收集、计算时同样有可能侵犯到当事人的隐私权、被遗忘权等。大数据时代对隐私权的侵犯已被学界多次讨论,因为算法技术的测算已经超越了人类对于自身信息的认知,这是对人隐私权的侵犯。智能司法对于证据的搜集、处理时应该考虑到对个人信息权的保护,与当事人身份不再相关、不再具有效用、不充分的信息等过时的信息不能作为司法审判的依据,应尊重当事人的被遗忘权。这种对个人信息权的保护应被设计进算法的技术之中,从程序上避免对当事人个人信息权的侵犯。不管是出于对审判公正性的要求,还是对当事人在证据收集时隐私权与被遗忘权的保护,都是对当事人人格权的保护。

#### 结语

通过对智能司法的伦理问题的研究发现,智能司法与司法伦理之间存在着紧密与紧张两种关系。首先,人工智能的发展有助于实现司法公正的伦理价值目标,本文论述了人工智能可以辅助司法实现:准确性、重复性、快速性、抉择性、无私性这五种特征。现

阶段人工智能还只能处于辅助司法的地位,这也是因为人工智能技术尚不具备人的学习、思考、应变等创造性能力。正是因为人工智能还处于一种工具状态,所以,一些程序化的运行模式会导致智能司法与司法伦理之间产生紧张关系,比如算法歧视不利于司法公正、违背司法伦理;智能司法对传统司法体系的冲击,因为司法案件的多样个性与人工智能程序化思维的矛盾是阻碍智能司法进一步参与司法的重要原因;机器会面临着软件程序与硬件部件出错损坏的风险,如果因为机器或者系统的故障影响到司法判决的公正是对当事人极大的不公,故对待智能司法应保持谨慎的态度。最重要的是人工智能不具备价值判断功能,这也是智能司法只能是人工司法的辅助工具的原因。本文论述了众多智能司法对司法伦理的正面作用,但是笔者看来,人工智能何时都只是辅助人类的工具。

智能司法与司法活动之间应该保持着,智能司法是一种辅助司法实现公平正义的技术,替代不了整个司法活动。智能司法应该秉承的理念应该是:第一,严格对智能司法所使用的人工智能技术或机器进行伦理审查,确保人工智能技术或机器不会发生有违伦理的问题产生才可使用。第二,坚持法官能动司法为主体,智能司法作为辅助工具,结合法官经验,追求符合价值判断的公正审判,不动摇法官的主导地位。第三,应该通过制定法律规范的手段规制算法歧视,智能司法所使用的算法技术需要被评估,符合伦理的算法技术才能被使用。同时算法技术应当被解释,当事人的知情权不可忽视。第四,应当从司法受理案件开始到审判结束充分保障当事人的人格权。在我们利用智能司法辅助判案实现公正时,需要思考智能司法的过程是否符合人们的伦理道德,若违背伦理道德而实现的裁判公正,就像违背了程序正义只追求实体正义的审判,是不被现代法治所认可的。故,智能司法应当作为符合伦理道德的一种技术,辅助法官实现司法的公正。

#### 注释:

① 此处的德性应指道德德性。

② 白银连环杀人案从1988-2002年14年间先后11人遇害,2004年被公安部命名为“8·05专案”,在案件的侦查的初期,技术的限制使得上百名刑侦专家都束手无策,甚至对比了上十万的指纹数据库。直到警方建立了一种针对精子DNA检测技术的数据库,通过数据的记录、分析,最终才锁定并抓获案犯。

③ 1978年,日本广岛一切割机器人突发异常将工人切割致死;1979年,美国福特公司位于密歇根州工厂中工业机器人的机器臂击中工人致死。因为机器人失控导致工人丧命的案例很多,本文不一一列举。

④ 马锡五审判方式是在抗日战争时期,马锡五任陕甘宁边区高等法院陇东分庭庭长时所创造的深入群众的审判方式。马锡五深入群众,了解案件经过,讲究审判与调解相结合,简化形式,方便百姓。马锡五审判方式深受当地百姓的接受与推崇,并在陕甘宁边区得到了推广。

## 参考文献:

- [1] Nils J. Nilsson, *Artificial Intelligence: A New Synthesis*, Morgan Kaufmann Publisher, Inc., 1998, p.1-2. 转引自郑曦.人工智能技术在司法裁判中的运用及规制[J].中外法学, 2020, 3.
- [2] 北京大学西方哲学教研室编.西方哲学原著选读(上册)[M].北京:商务印书馆, 1981.
- [3] [古希腊]亚里士多德.尼各马可伦理学[M].廖申白译.北京:商务印书馆, 2003.
- [4] 左卫民.从通用化走向专门化:反思中国司法人工智能的运用[J].法学论坛, 2020, 2.
- [5] 沈寨.个案正义视角下司法人工智能的功能与限度[J].济南大学学报(社会科学版), 2019, 4.
- [6] 罗洪洋, 陈雷.智慧法治的概念证成及形态定位[J].政法论丛, 2019, 2.
- [7] Stephen Marsland, *Machine Learning: An Algorithmic Perspective*, Boca Raton: Taylor & Francis Group, 2014, p4. 转引自蓝江.从碳基伦理到硅基伦理——人工智能时代的伦理学浅论[J].道德与文明, 2020, 5.
- [8] 郑智航, 徐昭曦.大数据时代算法歧视的法律规制与司法审查[J].比较法研究, 2019, 4.
- [9] 季卫东.人工智能时代的司法权之变[J].东方法学, 2018, 1.
- [10] [德]菲利普·黑克.利益法学[M].傅广宇译.北京:商务印书馆, 2016.
- [11] 孙那.人工智能的法律伦理建构[J].江西社会科学, 2019, 5.
- [12] Sandra G. Mayson. Bias In, Bias out. *Yale Law Journal*, Vol.128, 2019: 2218-2300. 转引自谢澍.人工智能如何“无偏见”地助力刑事司法——由“证据指引”转向“证明辅助”[J].法律科学(西北政法大学学报), 2020, 5.
- [13] 刘艳红.人工智能法学研究的反智化批判[J].东方法学, 2019, 5.
- [14] 孔祥俊.司法理念和裁判方法[M].北京:法律出版社, 2005.
- [15] [德]伽达默尔.真理与方法(下卷)[M].洪汉鼎译.上海:上海译文出版社, 1999.
- [16] [美]霍姆斯.普通法[M].冉昊, 姚中秋译.北京:中国政法大学出版社, 2006.
- [17] 马长山.AI法律、法律AI及“第三道路”[J].浙江社会科学, 2019, 12.
- [18] [德]拉伦茨.法学方法论[M].陈爱娥译.北京:商务印书馆, 2003.
- [19] 聂长建.法律语用推理的三元维度研究[J].政法论丛, 2019, 5.
- [20] 彭金治, 杜忠连.社会转型期能动司法的法治意蕴[J].黑龙江社会科学, 2015, 2.
- [21] 马靖云.智慧司法的难题及其破解[J].华东政法大学学报, 2019, 4.

## Ethical Problems in Intelligent Justice and Their Solutions

Luo Hongyang Li Xianglong

(Law School of Zhongnan University of Finance and law, Wuhan Hubei 430073)

**【Abstract】** The use of artificial intelligence in justice involves prominent ethical issues. Intelligent justice and judicial ethics are closely related and tense. The close relationship between intelligent justice and judicial ethics refers to the realization of judicial justice by intelligent justice. A very important factor of judicial justice is the acquisition of evidence which has great relevance to knowledge and wisdom. when artificial intelligence used in judicial, in terms of evidence acquisition, it has the following advantages over manual labor: accuracy, repeatability, speed, choice, and selflessness. The tension between smart justice and judicial ethics means that smart justice may do harm to judicial justice as following: the impact of algorithmic discrimination on judicial ethics, the changes of the nature of judicial justice and the dominant position of judges, the errors of big data and robots, and artificial intelligence itself has no value judgment. It is necessary to take precautions for the malignant risks of judicial ethics under the tensions and make plans: recognize the absence of the ethical in artificial intelligence justice, strengthen ethical review; control expectations of artificial intelligence justice, strengthen the dominant position of judges in the judicial system; regulate the algorithmic discrimination to protect the personality rights of the parties.

**【Key words】** intelligent justice; judicial ethics; judge-led; regulate the algorithmic discrimination

(责任编辑:唐艳秋)