欢迎按以下格式引用:陈欢,刘聪,齐浩宇.AIGC 技术辅助刑事审判的优势、风险与规制——以 ChatGPT 为例[J].长江大学学报(社会科学版),2024,47(4):113-118.

AIGC 技术辅助刑事审判的优势、风险与规制

——以 ChatGPT 为例

陈欢　　刘聪　　齐浩宇

(长江大学 法学院,湖北 荆州 434023)

摘　要:随着技术的进步,法律人工智能正在逐步从“非裁判”领域进入“裁判”环节,以规制法官的主观倾向,但受制于技术水平,传统法律人工智能不能满足司法实践需求。以 ChatGPT 为代表的 AIGC 技术,具有深度学习、反馈强化学习、强交互性等优势,能够为刑事审判带来深度技术赋能,然而,ChatGPT 类技术依旧难以摆脱自身存在的算法黑箱、知识幻觉、算法歧视、信息安全等问题,如何规避新兴信息技术赋能刑事审判过程中的风险,成为新时代智慧法院建设的重要课题之一。AIGC 技术与刑事审判的结合是技术要素对于公权力运行的介入,新司法生态下,相关风险的融贯性治理应当兼顾技术重塑与制度构架。

关键词:刑事审判;智慧法院;AIGC;ChatGPT

分类号:D925.2　**文献标识码**:A　**文章编号**:1673－1395(2024)04－0113－06

AIGC(Artificial Intelligence Generated Content),即人工生成内容,是指可以利用人工智能技术生成新内容的一种技术类型。ChatGPT 则是这一领域的现象级产品,是由 OpenAI 公司于 2022 年 11 月发布的大语言模型,很多研究者将其视作“引领第四次工业革命的关键性技术”。在法学学术领域,ChatGPT 为司法工作赋能,早已成为前沿命题,且 ChatGPT 的不断迭代仍在为该领域带来新的可能性。

在域外应用层面,英国上诉法院的 Birss 法官利用 ChatGPT 来总结其不熟悉的法律理论,并将结果运用于正式裁决。印度旁遮普邦和哈里亚纳邦高等法院的法官曾根据 GPT-4 出具的长达 94 页的报告内容作出拒绝保释的裁决。在这两个案例中,法官一致表示,ChatGPT 是“减轻审判压力的有力工具”。而我国在刑事审判“融智性”改革过程中,普遍出现了因技术壁垒导致的 AI 辅助审判实质性减负效果差、案件办理指引精度不足等问题,未能达到智慧法院改革所追求的“智能化审判”“同案同判”等审判目标。ChatGPT 类技术或可成为上述困境的破局之道。

一、AIGC 辅助刑事审判的优势

法律工作既然以自然语言为载体,相应的,人工智能就应当具备强大的自然语言处理能力,但就目前的发展情况来看,算力、算法、数据都不具备优势的传统法律人工智能并不能在此细分领域展现出良好的性能。ChatGPT 的横空出世让众多法律工作者看到了突破技术瓶颈的希望,对其抱以极大期待。笔者将在此部分尝试通过对比分析来探究 ChatGPT 的技术优势。

收稿日期:2024－05－18

基金项目:湖北省高等学校哲学社会科学研究重大项目“未成年人虚假供述影响因素体系化构建”(21ZD047);长江大学国家级大学生创新训练计划项目“生成式人工智能在司法实务中的应用研究”(202410489006)

第一作者简介:陈欢(1982—),女,湖南武冈人,副教授,博士,主要从事司法制度与诉讼法研究。

(一)我国传统法律人工智能路线

我国传统法律人工智能有两种较有代表性的技术路线:第一种是统计路线,即以历史相似案件的判决作为在办案件的参考;第二种是专家系统(Expert system)路线,把量刑规范化指导意见所确认的规则转化为量刑计算的公式,输入对应情节后即可得出对应的裁判结果。[1](P238) 这两种技术路线存在着明显差异。

在我国司法实践中,专家系统是传统法律人工智能发展的主要趋势。但专家系统也有缺陷,以国内应用较为成熟的上海"206"系统为例,71 个常涉罪名的证据标准构建,共耗费 400 余名刑事案件专家与众多技术人员一年零八个月,共形成校验点 12989 个。初期构建证据模型时,有关定罪量刑的要素点也是依靠人工选点标注。[2](P129) 目前用于智慧法院建设的人工智能系统大多只是封闭的专家系统,有赖于知识的人工输入,还不具备深度学习的能力。[3]

这一技术限制导致系统运作过度依靠人工干预,造成系统知识更新耗时长、维护成本大、偶然性风险过高等一系列问题。此外,我国不同地区部分刑案的认定标准存在差异,地区之间的经济与科技发展水平不同,导致主要依靠人工标注的专家系统目前仍然是选择性地被用于某些法律场景,并非全国推广使用。这加剧了我国司法水平的差异性。

(二)ChatGPT 的比较优势

ChatGPT 采取"预训练(Pre-training)+微调(Fine-tuning)"的运行范式。预训练基于无监督学习,通过使用来自多个领域的大量未标注的语料对模型进行训练,使模型能够自主进行深度学习并具备丰富的语言表达能力。随后,在少量标注数据上对模型进行进一步调整,使其适应特定任务或领域。

相较于我国传统法律人工智能的专家系统路线,ChatGPT 的技术优势主要体现为以下三点。

1.深度学习带来的知识储备与更新能力

现今刑事审判不仅需要考虑既有的法律规范,还要做到"类案必须检索",许多案件还要考虑不同的学术观点及其背后的哲学观念。每个个案都需要丰富的知识储备,而且这些知识处在不断更新之中,这就要求法律人工智能如法官般保持"终生学习",时刻更新自己的知识图谱。但依靠人工标注和搭建数据库的专家系统很难完成这一点,固有的封闭性和小语言模型的本质使其难以胜任庞大的知识存储和更新需求。

ChatGPT 的最新版本 GPT-4 是一个拥有 100 万亿级别参数的大模型。据学术界既有研究可知,深度神经网络的学习能力和模型的参数规模呈正相关。[4]并且,模型采用无监督学习,这是一种无须手动标注标签的机器学习方法。具备以上两点优势的 ChatGPT 能够在没有人类帮助的情况下时刻进行法律数据的学习,让模型能够随着法律制度演进不断进行自我更新,始终为法官办案提供正确指引。

2.人类反馈强化学习带来的机器审判"类人化"

刑事审判需要发挥人类的主观能动性,讲求"情理法"的圆融,而既有法律人工智能往往只能凭借知识库和推理机僵硬地完成决策任务,且法官迫于考核压力,不得完全拒绝使用 AI。这就可能导致刑事审判从过于依赖法官主观判断走向完全机械套用理论,让审判工作变为马克斯·韦伯所言的"自动售货机"。

ChatGPT 的核心技术是 InstructGPT,采用基于人类反馈的强化学习机制(RLHF)[5]。RLHF 是一种强化学习和人类反馈相结合的学习方法,模型不仅能够从环境中获得奖励信号,还可以从人类专家那里获得反馈信息来指导其行动。[6](P58) 如此一来,应用于刑事审判领域的模型可以在人类的调教下,与专业人员的认知、方法论、价值导向趋同。通过整合人类专家的偏好,模型在个案中模仿专业法官作出决策,最大限度地达到机器审判的"类人化"。

3.强语言能力带来的审判工作便利化开展

法律语言的意义往往有一定的"波段宽度",理解这些语言必须发挥主动性和创造性,不然很有可能陷入误解。加之法律语言多是非结构性文本,很难以统一的标准加以解构。在此情形下,传统范式上采用有监督学习、小语言模型的技术架构,事实上难以胜任法律领域自然语言处理技术的需求。[7]

ChatGPT 则有着业内公认的强大自然语言处理能力,具体体现为语言理解能力和交互性。语言理解能力体现为 ChatGPT 能够理解法律场景中复杂的语境和语义关系,例如,GPT-4 就曾以超过 90%考生的成绩通过了统一律师考试(UBE)。交互性则体现为在语言理解的基础上,模型可以整合案件情节提取、类案关键词检索、量刑偏离度对比等阶段性工作,以人机交互的方式进行辅助工作,减少法官在输入端的工作,并使法官能够在输出端快速获取模型对案件的整体分析,从而促进刑事审判工作的便利化开展。

二、AIGC 辅助刑事审判的风险

AIGC 辅助刑事审判在我国目前基本停留在学术讨论阶段。在世界范围内利用 ChatGPT 类技术进行审判的案例也是极少数,对此,学界既有支持者,也有反对的声音,但生成式人工智能进一步介入司法裁判,已是无法阻挡的潮流。[8]

鉴于 AIGC 辅助刑事审判的先例过少,难以对其风险进行实证研究,而价值、法理、情理等方面的风险则是"前人之述备矣",故笔者将从现有 AIGC 技术本身入手,分析两者结合后可能引发的风险。

(一)算法黑箱及其学术研究领域的扩大化

算法黑箱是指由于模型内部结构复杂、运行自主性较强且人工无法干预等因素,在模型训练与运行、输出结果等方面出现的"不可解释性"。算法黑箱问题的具体成因复杂多样,总体可以归结为以下四点:庞大的参数空间和灵活的神经网络结构导致的内部复杂性,算法非线性排列与高互动特征引起的算法逻辑的不确定性,基于数据驱动的算法运行行为的不稳定性,商业竞争需求和政策规定带来的算法保密性。由于以上障碍存在,人工智能模型的可解释性往往只针对于其研发人员。而随着深度学习技术的发展,如今主流的 AIGC 模型具备无监督学习能力,算法基于训练数据构建程序模型来模拟人类学习行为等智能活动,并不断利用经验数据来完善已有程序模型,改善自身性能。[9]这就意味着模型能自主完成从输入端的数据投入、训练到输出端的结果生成、优化改进的"全周期闭环"。这一情形下,即便是模型研发人员,也仅能理解底层代码所规定的部分,具体化的模型运作机理对于几乎所有人来说都有着一张"普罗透斯般的面孔"。

不可否认,算法黑箱作用于刑事审判后,可能会出现模型所作出的辅助行为影响司法透明度的现象,导致被告遭受不公正判决,甚至造成法院的司法权威性受损等一系列严重情况。但现阶段关于算法黑箱的讨论出现了将问题扩大化的倾向,具体体现为以下两点。

第一,"可解释性"的概念泛化。关于算法解释的技术,其经典的分类法为"内部解释"和"证明解释"。[10]前一种解释方法强调利用说明模型或中间规则等方法,为人们提供窥视算法运作逻辑的透明窗口;后一种方法强调模型对输出结果作出有针对性、有说服力的解释,以确保结果层面的正确。运用"可解释性"概念时,要区分技术人员主张的算法可解释性与实际应用领域的算法可解释性。[11]一般来讲,算法领域的专业人员要求的可解释性为第一种解释方法;算法用户由于"算法素养"的限制,往往采取第二种解释方法,要求算法能够"自圆其说"。而在相关讨论中,可解释性概念往往被简单泛化为第一种路线,这实际上是对算法设计者和使用者的为难。

第二,算法黑箱问题的偏狭化理解。人工智能技术层面的算法黑箱的确只针对算法本身,但广义的算法黑箱可以泛指任何不透明的决策过程。在计算法学领域,有研究者认为"法官的自由心证也是算法"。黑箱问题在刑事审判中一直存在,并非人工智能时代的独有产物。系统引入后,只不过是引发黑箱的主体由法官演变成人工智能算法。[12]因此,具备黑箱风险的 AIGC 模型不应在刑事审判领域被彻底否定,相关责任主体应当如监督法官般多角度建立风险防控机制,为 AIGC 技术赋能刑事审判提供具体方案。

(二)问题数据导致歧视与幻觉

算法歧视是算法控制者利用算法技术实施的,以算法决策为实现形式的歧视行为。[13]在排除商业领域人为设置歧视的情况后,算法歧视的根源在于供模型学习的数据样本存在局限或偏见,并且这种偏见最终会以模型决策的形式呈现给用户。例如,芝加哥法院的犯罪风险评估算法 COMPAS 曾被证实对黑人犯罪嫌疑人造成了系统性歧视。此案例中,算法所展现出的偏见实质上是当时司法机关对黑人偏见的映射,法院所作出的各类带有"隐性偏见"的判决以样本数据的方式被模型学习,模型依照既有数据作出的决策必然带有歧视成分。参照这一逻辑,我们也可以对 ChatGPT 所隐含的歧视进行一定程度的纾解。OpenAI 的相关研究表明,ChatGPT 的前身 GPT-3 的知识图谱中,犹太文化往往会与"精英"挂钩,而伊斯兰教却时常与"恐怖主义"伴随出现。这一偏见其实源于模型无监督学习的语料中,宗教偏见内容过多,模型在表达偏见后未能得到修正反馈,从而不断强化偏见。

知识幻觉或称幻觉(Hallucination),是指人工智能系统对自身的能力或知识状态存在误解,从而导致系统做出错误的推断或决策。幻觉的成因主要是模型在某个具体领域深入程度不足,所以在涉及专业领域的知识纵深时缺乏有效输出能力。幻觉问题在以"文字接龙"为输出形式的 ChatGPT 中体现得尤为突出,其在专业领域的表现时常被用户诟病为"一本正经地胡说八道"。例如,笔者曾就我国的

故意杀人案向 ChatGPT 提问,要求它提供具体案例,在其提供的 5 个指导性案例中,有 3 个案例出现了“张冠李戴”的罪名适用错误,这一表现可能为司法决策带来灾难。ChatGPT 幻觉的症结在于其通用性和缺乏外部知识库,作为一个以聊天为初始功能的模型,ChatGPT 用于预训练的数据来源众多,例如维基百科、新闻文章、各类书籍等,这造就了 ChatGPT 的无所不知,同时也为其进行知识的“胡乱串联”埋下了伏笔。而且,ChatGPT 并不会存储它的学习数据,因此,在用户需要专业知识时,ChatGPT 仅能凭借其对训练数据的“记忆”来生成答案,这毫无疑问会为刑事审判带来严重的负面效应。

综合以上分析,我们不难发现,ChatGPT 的算法歧视与知识幻觉的主要诱因都是其训练数据存在问题。数据本身存在的偏见和数据收集的局限会导致算法歧视,而数据的良莠不齐和缺乏外部知识库则会形成知识幻觉,所以,在许多技术资料中,两者被合并讨论。以此作为突破口,解决模型数据问题,将会成为 ChatGPT 进行技术优化的一大通路。

(三)信息安全与数据主权风险

ChatGPT 自发布之日起,围绕其信息保护层面的讨论与质疑便从未止歇。反观 ChatGPT 的自身表现,也并非如乐观主义者所设想的那般无懈可击。早在 2023 年 3 月 20 日,OpenAI 公司就宣布开源库出现错误,致部分用户与 ChatGPT 的聊天记录被泄露。紧随其后,意大利官方就于当地时间 3 月 31 日以 ChatGPT 涉嫌违反数据收集规则为由对其实行禁用。以上并非个例,在 2023 年 OpenAI 作为被告的诉讼情况分析中,对于隐私权的侵犯是一个重要的诉讼原因。[①]作为依靠海量数据驱动、缺乏信息分辨能力的大语言模型,ChatGPT 对个人信息的侵犯具有必然性。传统司法系统中,公民个人信息的保护主要依靠案卷管理制度和对办案人员的保密性要求,在此情形下,心怀不轨者难以染指公权力保护下的个人信息。而在引入 ChatGPT 类技术辅助后,一个收集了大量案件信息的语言模型必然会成为不法分子“围猎”的对象。这就使得模型本身一个不起眼的技术漏洞可能成为密集型网络攻击的突破口,从而引发社会个人信息危机。

个人权利会因技术介入而受到侵犯,国家主权也同样如此。现在具备广泛影响力的 AIGC 产品如 ChatGPT、Gemini、Claude 等,都出自美国头部科技公司。虽然我国也出现了文心一言、讯飞星火等性能不错的大模型,但在技术自主性方面仍有较大提升空间。目前国内产品的开发依照国外既有的技术路线,训练数据大多也以英语为主导,在西方国家先发优势的影响下,我国很难保证数据安全能够在这一赛道上得到足够保护,从而引发数据跨域流动、信息出境等方面的危险。而刑事审判领域在个人权益与社会秩序方面的特殊性让数据安全威胁在放大效应下愈发凸显,国家数据主权受到极大挑战。因此,出于传统国家安全的考虑,参与国家审判权力运行的技术产品必须完全由我国主导开发,且要在保证功能的前提下,实现训练数据的专门化。

三、AIGC 辅助刑事审判的风险规制措施

AIGC 技术与刑事审判的结合是技术要素对于公权力运行的介入,新司法生态下,相关风险的融贯性治理应当兼顾技术重塑与制度构架。一方面,嵌入刑事审判领域的 AIGC 技术应当朝着“技术中立性”前进,即模型在研发阶段就应当确保其含有最低限度的商业目的与潜在价值风险;另一方面,制度应当在尊重科技成果的基础上,利用相关规制手段最大限度防范化解司法风险,确保公共领域内算法的运行全过程都有相关责任主体进行监管。

(一)重申可解释性与加强顶层设计

在如今的大多数应用场景下,具备可解释性成为 AI 模型决策被用户采信的必要条件。但可解释性在面对不同的解释对象时,内涵应当有所区别。以 AI 领域专业人员的视角看待可解释性,模型的决策逻辑应当以机器语言展现出具有因果关系的推理决策能力,这种以算法解释算法的“解构性理解”方式技术门槛高,对于非专业用户不具备推广条件。作为一名普通用户,往往希望模型能用自然语言或简易逻辑图示的方法,对其输出结果进行“证明解释”。为应对法官“自由心证”这一“黑箱”,我们要求法官对裁判结果承担释明责任,以在裁判文书中进行充分说理的方式规制不透明决策。这一路径对于有着人机交互能力的 AIGC 模型同样适用,一个可解释的交互式人工智能可以针对特定使用者,在交

① 严嘉欢:《2023 年以来 OpenAI 公司作为被告的诉讼情况对比分析》,“清华大学智能法治研究院”公众号,2023 年 12 月 25 日。

互过程中提供细节和原因,使得系统背后的行动逻辑能够被用户所理解。[14]基于此原理,法官可以要求辅助刑事审判的 AIGC 模型对其输出结果进行强化说理,来窥视决策过程、降低黑箱风险。当然,AI 技术层面的可解释性也不容忽视。应用于高风险领域的模型结构应当适当简化,割舍部分非必要功能,以换取更高的可解释性。

在相关制度构建方面,我国应当以智慧法院建设的需求为着眼点,兼顾 AIGC 技术辅助下刑事审判领域的效率与公平,完善我国对于人工智能"黑箱"风险的监管体系,突出发展负责任的人工智能。

第一,建立统一的技术标准。我国传统法律人工智能系统往往由地方司法机关与国内头部互联网企业共同开发,如此一来,各地系统的决策水平、技术风险也都不尽相同,在制度层面难以统一把控。因此,国家应当针对刑事审判领域的 AIGC 技术制定统一的技术标准,充分发挥最高人民法院在智慧法院建设中的顶层设计作用,将"算法透明""可解释性"融入到统一技术标准中。同时,相关标准应当考虑我国地区之间经济、科技发展不平衡的客观现实,并保证省级行政区域内法院智能化程度的统一和均衡。在实现柔性监管的同时,确保被告人不会因审级变动而遭受智慧法院建设差异化带来的不利影响。

第二,健全算法监管制度。2023 年 8 月 15 日起施行的《生成式人工智能服务管理暂行办法》表明我国已开始以立法规制 AIGC 算法。下一步,我国应当制定更加完备的法律法规体系,出台更为细致的《人工智能法》,对人工智能行业进行系统规范,通过设立市场准入、增强运行监管、落实市场清退等方式全流程监管"黑箱",并将法律、金融、医学等高风险领域的人工智能以特别法加以规制。底层算法应当面向政府和第三方非商业监管机构进行开源,以方便事中监督和事后追责,并确保特殊领域不会因商业或其他因素而引发"黑箱"风险,最大限度地保护公民利益。

(二)建立专门数据库与公私协同监管

现阶段,ChatGPT 处理自然语言任务的本质是"使文本合理延续"。这里的"合理",更多是语义、语法层面的连贯和通畅,而知识准确、价值对齐则较少能被兼顾到。在此情形下,供其学习的数据很大程度上影响着它的表现。基于 ChatGPT 良好的泛化能力与鲁棒性,引入外部专家知识[15],成为其优化改进的良好方法。专门数据库的数据可以弥补其在垂直领域的不足,为其向特定领域发展提供可行路径。为刑事审判领域的 AIGC 模型提供一个专门的司法数据库,可以使其从中解构出正确价值导向,从而逐步消解"算法歧视",依靠对外部数据库的检索而非模型"记忆"对问题进行回答,以最大程度地避免"幻觉"。这一专门数据库应当囊括我国有关刑法和刑事诉讼法的各类法律语料以及裁判文书网、人民法院案例库所载的刑事案例。并且,这些数据应当经过严格质检,消除涉嫌歧视的内容,确保模型的每一个推断都具有高度盖然性。

落实监管主体对于国家系统性防范算法歧视和知识幻觉问题具有重要意义,在此方面,公权和私权应当协同促进,共同维护公民的个人权利不受算法风险的危害。

其一,在公权层面上,赋权监管主体对于辅助刑事审判的 AIGC 模型的数据进行全方位的监管,其内部数据流动的全过程,即从因法律现象而产生到因信息时效性被清除,应当由相关部门进行统一规范,进而通过数据控制来规制算法权力、减少算法歧视。[16]此外,我国应当构建核验机制,由核查人员在模型的输出端进行检验和控制,由输出端溯源输入端,并以此为依据对相关主体进行问责。对于模型内部算法,公法应当通过透明化、审批报备等手段来确保刑事审判领域不会出现"大数据杀熟"这类商业领域人为设置歧视的情况。在此方面,日本的《人工智能运营商指南(草案)》可以为我国提供借鉴,将人工智能研发者、提供者、使用者分立,确保人工智能全生命周期都能受到法律规制。

其二,在私权层面上,赋予公民足够的权利来对抗问题数据可能引起的错误裁判,强化公民面对算法的维权意识、风险辨识能力,增强公民主体意识。在此基础上,公权应当为私权行使做出必要铺垫,确保公民对于模型建议的知情权,建立申诉机构,细化纠错机制,保障公民独立人格不受侵犯。此外,应当发挥程序这一"社会制度化最重要的基石"的作用,将对于模型风险的规制融入到刑事审判程序。如果被告方以模型可能对其造成歧视为二审上诉理由,则二审不得有模型参与,防止实质性救济缺失。

(三)建立数据保护机制与推进技术自主

AIGC 模型应用于刑事审判后,其对个人信息的摄取具有必然性,个人信息的集中化又会令其成为公共数据的载体,建立数据保护机制对于保障信息安全具有必要性。《中华人民共和国数据安全法》和《中华人民共和国个人信息保护法》都将信息数据按其涉密性进行分类分级监管。刑事审判方面也可

以效仿此做法,对模型内公共数据进行分别存储并予以保护,例如,可以将案件信息按其法益侵害性、社会影响力、伦理争议性等因素进行分级,对于不同等级的数据予以不同程度的监管。对于高涉密性案件还可以通过数据清洗、数据标注等方法对数据进行"降敏",减少数据外泄可能带来的影响。此外,还应当提升模型抵御网络攻击的能力,以技术手段增强模型抗风险能力,通过系统内部优化,防止技术漏洞并建立长效网络攻击防范机制,第一时间对外部攻击进行识别并做出反应。GPT-4 就曾在 FAR AI 实验室进行的风险测试中暴露出安全漏洞,其 API(Application Programming Interface)会因过多良性样本的输入而对有害样本"放下戒备心",成为不法分子套取违法信息的捷径。因此,在正式投用前,技术提供方应当设立足够的"实验缓冲",重视暴露出的安全短板,有效加强后续抗风险能力,为信息安全设立坚实屏障。

我国在 AIGC 模型开发方面,必须有足够的技术自主性。对于刑事审判这类应用于高风险领域的模型,必须完全由我国掌握。应当以既有 AIGC 模型为借鉴,开发具有高技术自主性的国产专门模型。AIGC 模型的开发不仅包括程序的构建,还包括对初始模型的训练。我国于 2024 年 3 月 29 日在北京发布了第一个人工智能数据训练基地,该基地能够为国产大模型提供足够的算力支持,为我国的法律人工智能奠定技术自主的算力基石。在此基础上,应用于刑事审判的特殊垂类模型必须以中文法律数据进行训练,以防止外部强势文化和非法律语料造成刑事审判内在价值秩序的含混。在数据服务监管层面,相关服务器必须由国家机关掌握,最大限度地防止数据外流,保障国家数据主权。

此外,模型的维护和迭代也必须依赖技术专家,如果继续走"技术外包"路线,模型便很难排除商业性质。刑事审判等公共领域的 AIGC 模型有必要建立专门的技术人员团队,具备"法学+人工智能"的交叉学科知识,专门负责模型开发后的人工标注、指令微调、维护与升级等工程,为国家司法智能化进程提供人才保障。

四、结语

AIGC 领域的变革正以我们难以预料的速度上演着,作为一种新事物,AIGC 出现缺陷和短板在所难免,我们应当以更加开放的姿态去迎接机会、规避风险。上文所述各种风险并不是独立的孤岛,而是彼此联结、互相影响的共同体,相应的规制手段也不是孤立片面的,而是通用性与针对性的统一。只有顺应人工智能时代的浪潮,我国的刑事审判才能回应时代需要;只有做到制度与技术的协同治理、同向发力,才能为我国的智慧法院建设进程保驾护航。新时代智慧法院建设将促进人类的价值正义同人工智能的数字正义在刑事个案上融合,以算法的绝对理性斧正审判人员的内心确信,共同推进我国的刑事审判朝着统一化、智能化的方向发展,让每一位被告人都能感受到技术变革下公正、高效的新司法生态。

参考文献:

[1]姜伟.法律人工智能导论[M].北京:北京大学出版社,2023.

[2]崔亚东.人工智能与司法现代化[M].上海:上海人民出版社,2019.

[3]郑戈.大数据、人工智能与法律职业的未来[J].检察风云,2018(4).

[4]朱光辉,王喜文.ChatGPT 的运行模式、关键技术及未来图景[J].新疆师范大学学报(哲学社会科学版),2023(4).

[5]钱力,刘熠,张智雄,等.ChatGPT 的技术基础分析[J].数据分析与知识发现,2023(3).

[6]范煜.人工智能与 ChatGPT[M].北京:清华大学出版社,2023.

[7]王禄生.ChatGPT 类技术:法律人工智能的改进者还是颠覆者?[J].政法论坛,2023(4).

[8]郑曦.生成式人工智能在司法中的运用:前景、风险与规制[J].中国应用法学,2023(4).

[9]谭九生,范晓韵.算法"黑箱"的成因、风险及其治理[J].湖南科技大学学报(社会科学版),2020(6).

[10]周翔.算法可解释性:一个技术概念的规范研究价值[J].比较法研究,2023(3).

[11]赵泽睿.算法论证程序的意义——对法律规制算法的另一种思考[J].中国政法大学学报,2023(1).

[12]吕泽华,渠澄.刑事审判"融智性"改革的现实困境及进路探析[J/OL].长白学刊,2024:1-13[2024-03-20].https://link.cnki.net/urlid/22.1009.d.20240117.1029.002.

[13]宁园.算法歧视的认定标准[J].武汉大学学报(哲学社会科学版),2022(6).

[14]吴丹,孙国烨.迈向可解释的交互式人工智能:动因、途径及研究趋势[J].武汉大学学报(哲学社会科学版),2021(5).

[15]秦涛,杜尚恒,常元元,王晨旭.ChatGPT 的工作原理、关键技术及未来发展趋势[J].西安交通大学学报,2024(1).

[16]肖东梅."后真相"背后的算法权力及其公法规制路径[J].行政法学研究,2020(4).

责任编辑 叶利荣 E-mail:yelirong@126.com