

人工智能法律服务中的算法偏见与规制路径

林厚任, 韩东爽*

三亚学院法学院 海南三亚

【摘要】人工智能在法律服务领域前景佳、潜力大,应用 AI 的律所营收增长,凸显技术重要性。我国人工智能立法和监管框架渐完善,司法部门需出台内部规定防范风险,保障司法公正、数据安全与权威。人工智能存在算法偏见,数据利用要正当合法,保障隐私。隐私保护的算法歧视隐蔽复杂,现有法律界定与协调不足。司法领域算法偏见有身份歧视、数据偏差等,影响司法公平,治理需动态监管体系,但面临技术迭代等障碍。为此,提出加强数据安全、明确责任主体、加强动态监督,实现法律与技术协作,构建源头防线,确保数据合规,厂商与机构共同审查,转向平时监管,纳入系统,紧抓源头,构建体系。所以人工智能在法律服务中的应用、偏见与路径相互影响,发展需规则与技术共进,重视监督,减少算法问题,保障数据合法透明,法律与服务机构保障数据安全、提供公正服务。

【关键词】人工智能;人工智能法律服务;算法偏见;算法黑箱;大数据

【基金项目】2023 年大学生创新创业训练计划项目(项目编号:S202313892003):利用人工智能和自然语言处理改进和优化法律服务;2023 年海南省哲学社会科学规划课题(HNSK(ZC)23-193):基于陆海统筹的海南生物多样性法律保护研究

【收稿日期】2025 年 12 月 9 日

【出刊日期】2026 年 1 月 6 日

【DOI】10.12208/j.ssr.20260006

Algorithmic bias and regulatory path in artificial intelligence legal services

Houren Lin, Dongshuang Han*

Law School of Sanya University, Sanya, Hainan

【Abstract】Artificial intelligence has a promising future and great in the field of legal services, and the increase in revenue for law firms using AI highlights the importance of technology. China's AI legislation and regulatory framework are gradually improving, and judicial sector needs to issue internal regulations to prevent risks and ensure judicial fairness, data security, and authority. Artificial intelligence has algorithmic bias, and data utilization must be legitimate and legal protect privacy. Algorithmic discrimination in privacy protection is concealed and complex, and existing legal definitions and coordination are insufficient. Algorithmic bias in the judicial field includes identity discrimination and data, affecting judicial fairness, and governance requires a dynamic regulatory system, but it faces obstacles such as technology iteration. To this end, it is proposed to strengthen data security, clarify the entity, and strengthen dynamic supervision to achieve collaboration between law and technology, build a source defense line, ensure data compliance, and jointly review by manufacturers and institutions, turning to routine supervision incorporating into the system, focusing on the source, and building a system. Therefore, the application, bias, and path of artificial intelligence in legal services influence each other, and development to advance with rules and technology, pay attention to supervision, reduce algorithmic issues, and ensure legal and transparent data. Law and service institutions ensure data security and provide fair services.

【Keywords】Artificial Intelligence; AI legal services; Algorithmic bias; Algorithmic black box; Big data

作者简介:林厚任,男,从事人工智能法学领域的研究;

*通讯作者:韩东爽,女,三亚学院法学院讲师,海南自贸港海洋资源法治研究院助理研究员,从事环境与资源保护法学、人工智能法学领域研究。

人工智能正在深度重塑法律服务业态,从智能合同审查、类案检索到法律风险预测,其高效性与低成本优势显著提升了法律服务的可及性,人工智能服务也能帮助法律制定部门进行法律制定以及提供生活实践素材。然而, AI 法律服务的核心是算法,非绝对的客观,其训练数据的偏差、算法模型的“黑箱”特征及开发者的隐形偏好,易催生算法偏见^[1],可能会导致法律决策不公、权利救济失衡等问题,例如对特定群体的法律需求产生法律制定系统上的排斥。这就表达了当今社会要揭示算法偏见的潜在风险与现实危害,要坚持构建科学、可行法律规制路径的必要性与紧迫性,所以平衡技术创新与法律正义成为关键命题。

1 人工智能在法律服务中的应用现状

当前,人工智能在法律服务领域已经进入快速发展的阶段,它正在重塑传统的法律服务,推动着传统法律服务向着高效便捷方向发展,也推动着法律服务向智能时代发展^[2]。一些传统法律专业人士对 AI 的态度也是从排斥使用变成谨慎使用再变成积极使用,所以现在人工智能在法律服务中是规模化使用,呈现模块化,这也是当今社会的大势所趋。但是人工智能的出现是机遇与挑战并存,人工智能在法律条文的研究以及法律文件的审阅中展现出的高效性和条理性,但是人工智能在案件文件中并不能体现出传统审阅中受到的人文关怀,社会舆情等影响,只是一味地总结经验,忽略了人在社会中发展的影响。当今的司法部门积极拥抱和规模化应用人工智能服务,在法律文书起草中展现出高效便捷,能将耗时极长的工作进行检索压缩至较短的时间完成,并且能确保文件的格式规范以及引用文献正确^[3],但是人工智能可能会生成与实际不符合且不准确甚至是虚构的信息,这就需要专业的司法部门专业人士进行审查和辨别,并且确保工作符合职业规范。

人工智能的发展依赖于科学技术的发展,科学技术的发展需要社会发展去支撑,需要资金链的支持, AI 法律服务前景广阔,发展潜力高,所以带来了发展技术的驱动力。应用 AI 技术的律所也实现了显著的业绩增长,有报告指出,这类“成长型律所”的近四年营业收入增长了近一倍,而客户和案件数量呈现上涨明显,实现了高效的工作效率,这表现出技术在法律服务中的重要性。当前,人工智能相关的立法和监管框架也在逐渐发展和完善,我国通过修订《网络安全法》加入了 AI 安全发展的框架,并且全国人大常委会已将人工智能法列为预备审议项目,为未来的人工智能发展奠定更

坚实的基础,中国也在国际上积极推动全球人工智能监管框架的建立,我国的司法部门运用 AI 带来的便利性,也应该出台相关部门内部的使用规定,确保相关部门人员使用人工智能工具时的合理性,保证司法系统的高效运转,帮助司法系统工作人员的工作效率,制定其核心在于保证司法系统的公正司法和司法数据的机密,维护司法部门的权威,并且要时刻保持警惕,防范未知的风险。

2 人工智能在法律服务中的算法偏见

2.1 数据利用的正当性

人类社会正加速迈入一个由数据驱动的智能时代。数据不仅正在深刻地改变我们的生产方式与生活习惯,然而在带来巨大的便利同时,也逐渐暴露出算法歧视这一亟待关注的社会问题。数据利用要保证数据的合法性,符合《数据安全法》等法律法规的规定,数据利用的目的要保证其正当性,保证其数据的隐私安全,数据要在合理合规的途径下进行使用,不得滥用,避免数据通过不正当途径流传,也要保证数据来源的合法性,也要避免冗余收集,要确保他人对数据有知情权和控制权。例如,最高人民法院指导性案例 264 号^[4],某钢铁有限公司诉某电子商务股份有限公司侵权责任纠纷案中,某电子商务股份有限公司依法采集企业数据,经符合有关标准的编制方法加工形成数据产品并合理利用,未对企业权益造成损害,法院明确,在这种情况下,相关企业要求数据处理者承担侵权责任的,人民法院依法不予支持。该案例体现了数据处理者在合法合规采集和利用数据的前提下,其数据利用行为的正当性受到法律保护。

2.2 隐私保护的必要性

与传统歧视不同,大数据时代的算法歧视呈现出隐蔽性、复杂性,系统性和不可逆性四个主要特征^[1]。算法歧视往往隐匿于自动化决策过程之中,难以被个体或公众察觉,人工智能的算法技术依赖于数据训练,人们将大量数据导入人工智能,利用快速且便捷的计算,将这些大量的数据总结出共同的特性,这就是“大数据杀熟”^[5],当今社会大量使用人工智能进行分析,导致了用户数据处于公开且透明的状态,缺乏隐私保护,由于法律规则的体系性缺失。现有法律框架多以原则性条款为主,如《个人信息保护法》虽规定“不得利用算法实施不合理的差别待遇”,但缺乏对算法歧视的具体界定标准。这种立法模糊性导致司法实践中出现“同案不同判”现象:某招聘平台算法歧视案中,法院因无法认定“不合理差别”的量化标准而驳回诉求;另

一信贷算法歧视,因法官对“算法透明度”的理解差异产生不同裁判结果^[6]。此外,跨部门立法协调不足加剧了制度碎片化,数据保护法、反不正当竞争法、就业促进法等对算法歧视的规制存在交叉与空白,形成“都管都不管”的监管真空。

2.3 司法领域的应用性

分析人工智能的算法偏见在司法领域的应用性的体现应该从分析和治理去描述。在司法领域运用人工智能进行分析和解答已经成为常态化,但是人工智能在司法领域提供的帮助是一把双刃剑,甚至有一些算法偏见已经对司法领域构成了实质性的威胁,算法偏见有各种表现形式,从身份歧视到数据偏差,这就严重地影响了司法领域的公平正义。身份歧视与偏见:算法系统基于特定身份特征(如种族、性别、地域等)进行歧视性判断的现象在司法领域尤为突出。2025年美国一起引人注目的诉讼案件显示,Way mo公司的AI身份验证系统将一名名叫Nasser Mohamed的医生错误地识别为恐怖分子,仅仅因为他的名字听起来像中东或穆斯林名字,这就体现出人工智能在算法偏见上有自己选择性,并且依据自己的选择进行执行。数据存在偏见和偏差,现行人工智能所给司法领域提供的观点大部分来源于历史的裁决,而这些历史的判决可能因为当时法律的不完善和错误的评判,以及历史性地不公平的判决^[7],而这些外在的影响固化了人工智能的判断,导致人工智能在给司法领域提供服务的时候会给出与历史相似的答案,而这些答案是不符合当今司法领域的判决。所以当人工智能固化了这种学习思维方式,这样的人工智能会成为阻挡司法进步的力量。并且人工智能的算法黑箱与透明度缺失使司法领域存在逻辑不通的问题,严重影响了当事人对司法判决的认知和对司法公平正义的质疑。

算法偏差在司法领域的治理需要依靠动态监管体系的建设^[8],但是动态监管体系的建设难度较大,需要大量的人力物力资源的堆砌,所以遭到的障碍较多,动态监管体系的建设遭遇三重障碍:首先,技术迭代速度远超监管响应能力。GPT-4等生成式AI的出现使算法歧视形式从“数据偏见”升级为“内容操纵”,而现行监管规则仍停留在传统决策算法层面。其次,跨平台数据流动加剧监管难度。某社交平台算法通过用户跨平台行为数据实施歧视性推送,但监管部门因数据主权限制无法获取第三方平台数据导致证据链断裂。最后,监管资源与算法复杂度严重失衡。某省级网信办仅有5名技术人员负责审核辖区内237个算法系统,平均每

人需监管47.4个复杂AI模型,监管效能可想而知。更值得关注的是,动态监测技术本身存在伦理风险,某监管部门使用的算法监测工具因设计缺陷,误将正常商业策略识别为歧视行为,引发企业集体诉讼。所以动态监管体系的建设需要进行大量的验证和实践考察,而不是一蹴而就的。

3 人工智能在法律服务中的规制路径

3.1 加强数据安全

数据安全是当今数据技术领域的热点问题,它的核心内容是保护用户的隐私、完整性、可用性等^[9]。保证数据不会为非授权人员和机构访问,对于维护个人隐私权和企业商业机密至关重要。结合上面两个板块去分析,如何加强数据安全需要技术人员建设可靠的、可信的基础设施,通过建设可靠的基础设施保证传输、分析等过程数据的保密性和完整性,法律服务的人员应当起到监督的作用,保证数据不被人为泄密和窃取,人工智能提供法律服务依靠于硬件的输出,也要保证硬件的安全和安全操作的环境。加强数据安全要有完善的法律监督和检察,要保证数据的公有化和私有化,保证数据在设定的网络内传输和使用,防止泄密,也要从访问端口去控制,保证最终控制权在可靠可信的人群,防止不法人员和带有恶意的人员进行篡改,并且请专业的网络人员,对一些重要的数据和端口进行人为加密,构建加密数据传输通道,保证数据传输的安全。

立法部门也应该制定相关的法律为数据安全保驾护航,避免被不法分子利用法律漏洞进行数据的篡改等违法行为。加强数据安全需要设定访问权的界限,减少算法黑箱的干扰,尽量避免数据的泄露,并且根据数据的重要等级来划分访问的差别等级,也要保证人工智能在管理数据上不会生成看似合理,实际逻辑错误的的数据,要对人工智能管控的数据进行人工审核,人工智能提供的法律服务需要法律实践者和技术团队的双重协作,筑牢数据安全的红线。

3.2 明确责任主体

结合上一板块来看,这个板块讨论的问题是由技术支持转变为权责规范,人工智能在法律层次提供的的数据推动了法律服务事业的发展,但是提供的的数据也是一个。明确责任主体应该有多重定位,而第一个定位就是第一道防线的构建以及人工智能服务提供的源头责任,人工智能提供的法律服务所生成的数据应该确保在一定范围内运行,保证这些数据符合现行法律的规范,具备合法性,并且源头设计者要对一些敏感数据进行脱敏训练,保证提供的数据合理合规,如果人工智

能提供的数据过于低效, 必然会导致法律服务的质量下降, 甚至会承担相应的法律风险。

人工智能提供的法律服务是依靠算法来呈现的, 而呈现的结果需要法律服务机构和人工智能厂商来共同审查, 不能交给单个一方来审查, 人工智能厂商要向法律服务机构提供一定的技术透露, 并且如实提供算法能力、偏见性的报告, 以及研发的人工智能的局限性体现在何处, 而法律服务机构使用人工智能提供的数据, 不能以算法黑箱的数据方面问题来推卸自己的责任, 法律服务机构应当对数据进行多次审查, 保证数据的真实性和准确性, 最终的审查应该是人为的查看, 这是否符合社会主流价值观、是否合法合规, 这就需要工作在一线法律的工作人员来进行审查了, 所以在明确责任主体, 应当分析不同阶段, 通过精细化的划分, 确立不同的责任主体。

3.3 加强动态监督

如何加强数据安全和明确责任主体, 正是需要加强动态监督, 这就意味着不再是事后补救, 而是重在平时的周期性发展, 需要纳入人工智能的系统性工作。动态安全是指以承认维护和促进数据的利用为前提, 通过规范数据处理活动合理控制风险, 维护数据安全。动态安全范式的技术特点是安全软件会监控所有系统软件的各种异常行为和监控系统文件的各种异常行动。动态的监管, 是为了保证人工智能提供的法律服务可以保证公平性和合法性, 从而搭建可靠的人工智能法律服务体系, 更好地服务于全社会。如何监管人工智能, 需要结合上面文章的出发点, 人工智能提供的任何服务都是依靠算法, 但是有些数据给人工智能提供的依据使得算法出现了偏见和歧视, 所以动态监管就是需要将这些歧视和偏见尽可能降低, 在最大程度保证法律服务的数据公平公正, 动态监督最重要的应该是紧抓源头, 随时可以溯源, 保证源头数据不可篡改。也应当学习书籍出版的程序, 不断的校验, 依靠多道程序来保证数据的偏差和歧视不断减小, 符合现在社会的主流价值观, 动态监督需要保证法律制度和技术协作的携手共进, 要构建一个实时感知、精准定位, 快速修复的动态监督体系, 动态监督也是人工智能在提供法律服务中的坚实底座。

4 结语

人工智能在法律服务中的应用现状、算法偏见、实践路径是相互搭建和影响的, 人工智能不断发展完善, 离不开规则的制定和技术的支持, 二者是携手共进的关系, 要不断地调整使得人工智能提供的法律服务更加符合社会的发展, 也要重视动态监督, 减少算法黑箱的出现, 以及尽可能减少算法的偏差, 保证数据的合法合规和透明性, 有关法律部门和法律服务机构也应当保证数据在内部渠道传输和审阅, 做到不泄密、不篡改、不伪造, 保证为社会提供的法律服务公平公正, 打造一个社会可以依赖的人工智能时代。

参考文献

- [1] 左珊珊. 算法歧视的法律规制研究[D]. 内蒙古科技大学, 2025:12.
- [2] 龚善要. 生成式人工智能赋能公共法律服务的范式转型[J]. 东方法学, 2024(9):88-100.
- [3] <https://www.court.gov.cn/zixun/xiangqing/474471.html>
- [4] <https://www.ciplawyer.cn/articles/133396.html>
- [5] 王玲, 张明略. 人工智能赋能公共法律服务城乡均衡发展的困境与法治进路[J]. 山东理工大学学报(社会科学版), 2025(3):27-36.
- [6] 赵鑫, 邵奇, 李东艳. 人工智能技术在智能化法律服务中的应用[J]. 信息通信技术, 2019(2):62-66.
- [7] 岳平, 苗越. 社会治理: 人工智能时代算法偏见的问题与规制[J]. 上海大学学报(社会科学版), 2021(11):1-11.
- [8] 张怡. 人工智能时代的风险及法律应对[J]. 辽宁公安司法管理干部学院学报, 2020(2):75-80.
- [9] 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[J]. 山东大学学报(哲学社会科学版) 2019(5): 115-123.

版权声明: ©2026 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<http://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS