

# 生成式人工智能赋能司法裁判：风险阻力与应对策略

史静蕾

新疆大学政治与公共管理学院(国家安全学院), 新疆 乌鲁木齐

收稿日期: 2025年4月2日; 录用日期: 2025年5月13日; 发布日期: 2025年5月26日

## 摘要

研究探讨了数字化时代下生成式人工智能在司法裁判中的前景与风险, 并提出通过技术赋能, 生成式人工智能可在证据展示、文书制作、裁判辅助等方面极大提升司法效率和透明度。当前, 针对技术的深入应用存在内容可控性与真实性、裁判逻辑、以及司法伦理和监管保障等系列问题, 研究提出“辅助为主、审查为要”的应用模式, 强调人工智能技术与司法经验的协作, 以确保法律裁判的可控性与公正性。

## 关键词

ChatGPT, 生成式人工智能, 司法裁判, 智慧司法

# Generative Artificial Intelligence Empowers Judicial Adjudication: Risk Resistance and Strategy

Jinglei Shi

School of Politics and Public Administration (School of National Security), Xinjiang University, Urumqi Xinjiang

Received: Apr. 2<sup>nd</sup>, 2025; accepted: May 13<sup>th</sup>, 2025; published: May 26<sup>th</sup>, 2025

## Abstract

The study explores the prospects and risks of generative artificial intelligence in judicial adjudication in the digital era, and proposes that through technological empowerment, generative artificial intelligence can greatly enhance judicial efficiency and transparency in the areas of evidence presentation, document production, and adjudication assistance. However, the in-depth application of the technology still has a series of problems such as content controllability and authenticity, adjudication

logic, as well as judicial ethics and regulatory safeguards. In order to address these challenges, the study proposes an application model of “assistance as the mainstay, review as the mainstay”, which emphasizes the collaboration between artificial intelligence technology and judicial experience, in order to ensure the controllability and fairness of legal adjudication.

## Keywords

ChatGPT, Generative Artificial Intelligence, Judicial Adjudication, Intelligent Justice

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

## 1. 引言

近年来, DeepSeek、GPT-4.5 等生成式人工智能基于多模态数据输入, 通过深度学习与模型构建等技术, 具备强大的文本学习和理解功能, 可创造性地生成内容, 现已广泛应用于各领域。对此, 法学界或关注生成式人工智能训练数据的合理使用, 寻求构建更为公平的利益分配机制[1], 或聚焦生成式人工智能刑事风险的高效治理, 探讨如何在技术提供者、服务提供者、产品提供者等主体之间合理分配法律责任, 并实现人工智能治理的更新、转型乃至迭代[2], 甚至进一步思考生成式人工智能对司法审判带来的机遇与挑战[3]。

## 2. 生成式人工智能在司法裁判中的运用前景

### 2.1. 技术赋能与司法革新的时代需求

数字时代为司法活动提供了强劲的技术驱动力, 5G、人机交互、区块链、大数据以及以 ChatGPT、Deepfake、元宇宙为代表的深度合成技术蓬勃涌现, 大幅拓宽了司法的“生产力”边界。此前, 传统司法既要满足不断增多的纠纷解决需求, 又要应对诉讼资源、时间与空间等维度限制, 而生成式人工智能可利用算法、大数据来统一证据标准、简化或加速办案流程, 实现对信息的高度整合、自动检索与生成。一方面, 生成式人工智能通过不断完善的算法模型和对海量案件数据的深度学习, 可迅速解析控辩双方证据资料、识别要点并给出相关法律检索结果, 最大程度地支撑司法革新。近日, 多伦多大学联合硅谷科技公司研发的“Justice-1” AI 系统通过分析近 200 年全球判例、数百万条法律条文及社会舆情数据, 构建出动态裁决模型, 在庭审中, 仅用 12 秒便完成了双方陈词、证据及法条交叉对比, 并输出了 83 页的判决书, 支持 AI 司法的群体认为这将终结“同案不同判的历史顽疾”。另一方面, 数字化审判愈趋普及, “智能”裁判工具逐渐成为判案辅助的重要组成, 满足司法现代化深度革新的时代需求, 成为推动司法裁判数字转型的原动力[4]。随着中国式现代化数字法院建设, 案卷材料已实现电子化存储, 依靠光学字符识别等技术, 立案庭法官可对电子材料进行检索, 因此, 法院将电子材料输入生成式人工智能, 对其进行模型训练后, 生成式人工智能可根据当事人表述, 迅速作出案由、诉求判断。

### 2.2. 可视化呈现与审判程序的全新表达

传统示意证据在部分案件中发挥直观说明与辅助解释的作用, 在具体应用时受制于技术手段落后、制作过程繁琐。而生成式人工智能则基于海量数据与自然语言处理技术可实现自动输出文本、图形、音视频乃至三维模型, 为案件审理提供可视化“示意证据”。法官、检察官以及辩护人通过视听材料、空间

模拟技术与生成式算法的深度融合，更为直观地比照案件发生过程、分析相关要件，从而在庭审中获得对事实的更具象理解(即还原事实)。

在诉讼文书制作层面，生成式人工智能通过分析法律规定、历史判例对证据进行深度整合，协助或部分替代人工进行文书撰写，进而辅助法官精确提炼案件争议焦点、拟定判决文本结构；帮助律师准确定位检索条文、推送相似案例；对控辩双方所提证据一并进行逻辑梳理并生成初步分析[5]。无论是可视化方式还是文本生成模式均使司法活动以数字形式实现过程透明和高效衔接，让原先深奥的法律推理更易于诉讼参与人和社会公众理解，进而提升裁判过程的正当性与公信力。

### 2.3. 裁判方式的变革与恣意风险的约束

生成式人工智能在司法中的深度介入是对证据展示或文书制作的“工具性”优化，更可能带来对裁判方式本身的重塑乃至变革。其一，人工智能技术的介入可在一定程度上实现对不同审判者因经验差异、知识背景或情绪偏好而导致裁判结果不统一的校正，从而在统一尺度与客观标准的维度上发挥积极作用。其二，由于生成式人工智能善于从大量数据中提炼共通要素，并通过概率模型提供预测性输出，法官在决策时或将倾向于咨询或参考它给出的建议。哥伦比亚、印度的个案也表明，法官利用 ChatGPT 等生成式人工智能的意见来印证自我判断或直接纳入裁决理由，让“智能系统”在一定程度上获得了分享裁判权的话语空间。

相对而言，技术能辅助审理过程，却无法充当或替代人类的评判经验和价值判断。裁判方式对恣意风险的深层冲击依然值得关注：若过度倚赖算法而忽视个案中应保留的情感、伦理及主观能动性将使司法陷入形式化推理的泥淖，不利于充分保障当事人的权益。同等重要的是，生成式模型对自身训练数据的依赖也可能因先例不足、语料偏颇而输出倾向失衡，进而对司法公正构成新的挑战。在技术理性和司法经验理性的双重约束下，可严格限定与有效监管生成式人工智能应用场景，真正让人工智能在司法裁判中避免恣意裁判的滋生，实现对公平正义更为精细化的保障。

## 3. 生成式人工智能在司法裁判中的技术障碍

### 3.1. 生成式内容的可控性与真实性难题

在司法领域中，生成式人工智能基于多模态数据输入(如文本、图像、音频、视频等)进行深度学习与模型构建的能力，及其“创造性”的内容生成特性为审判实践提供了诸如类案分析、证据呈现和法律建议等辅助功能。这一技术性优势背后暗含着深刻的伦理和技术问题——尤其是生成内容的可控性与真实性。在数据输入环节，训练样本质量、来源、完整性及其涵盖的价值观念和情感倾向直接影响生成内容的准确性与可信性。若原始数据中存在偏差或包含虚假的证据样本、不准确的法律推理，系统更倾向于生成与真实情况严重背离的“示意证据”，对案件事实的认定及法律适用构成误导[6]。

除此之外生成式人工智能的深度合成技术可模拟真实场景，以高度逼真的方式展示证据和再现案发过程，高拟真度内容尽管在某些情况下可以弥补传统证据的不足，但也使法官、律师和其他司法参与者对其内容真实性产生疑虑，如生成内容未经严格审查将破坏司法裁判的正当性，影响公众对智能司法系统以及裁判结果的信任。

### 3.2. 算法推理对裁判逻辑的冲击与恣意风险

生成式人工智能可以通过检索海量法律规定、案例、证据材料，进而迅速给出判断建议或裁判预测，为法院统一办案标准、提高度量衡的客观性提供了技术支持。但大规模的概率计算与预测并不天然具备“法律适用”的准确性与正当性，算法在判断过程中无法完全重现三段论的演绎逻辑，也无法充分考量

个案所需的价值衡量。当法官过于依赖此类“快速决策”功能，或者在案件审理初步阶段即被算法建议所诱导，难免忽视人类经验、社会伦理等非量化考量因素，从而萌生将部分裁判权力让渡给机器的隐蔽趋势。正因如此，若缺乏对应的监督和规范，生成式人工智能在特定情形下对法官裁量空间的过度干预，可能会产生“结果同质化”或“机械化裁判”的弊病。更为严重的是，一旦系统面对训练数据缺陷或审判案例的多元局面时，错误或偏颇的判断建议会直接诱发恣意风险，反而妨碍了司法对公平正义的实质守护。

### 3.3. 司法伦理与监管保障的不足

生成式人工智能技术在司法领域的应用预示法律裁判的数字化与智能化转型，但同时也暴露出法律和伦理层面尚未完善的规范架构，存在对个人隐私、算法透明、法律责任分配等多重风险的潜在冲击。

生成式人工智能直接参与裁判工作的最大隐患之一是对程序正义的潜在威胁，即法官将人工智能生成的建议或分析结果直接作为裁决依据时缺乏科学严谨的审查机制，导致裁判阶段的透明度和公正性受到质疑。其生成的裁判建议由于拥有数据驱动的表面“客观性”，使法官过于信任和依赖而忽视自身的独立判断，无意识地依赖于生成式人工智能的数据分析和逻辑生成功能，忽略案件中情感考量、伦理冲突和社会公平的价值判断，进而抑制其对案件人物关系、社会背景等非量化因素的敏锐关注，使裁判逐渐被局限在形式化推理的框架内[7]。

这一问题的本质在于生成式人工智能依据以技术理性为主导的大数据和算法进行决策，其“理性”基于训练数据的质量及设计者的意图，决策过程是否会忽视法律实践中的重要伦理问题仍旧存疑。而裁判权则承载着宪法意义上的国家权威和法官的独立判断，涉及对案件事实与价值的衡量，而非单纯的理性分析。若将人工智能生成的结果直接作为裁判依据无疑将赋予技术以部分“裁判者”的身份，势必模糊传统法律中裁判主体与技术辅助之间的界限，引发对司法裁量权是否被技术稀释的疑问。并且，生成式人工智能依赖于训练数据的质量及偏向性，假使数据本就带有社会中存在的不平等和歧视性(种族、性别偏见或者阶层隔离等)，若算法无意中放大不平等并将其体现在裁判建议或示意证据中，司法活动可能从根源上失衡。

## 4. 重塑生成式人工智能在司法裁判中的应用模式

### 4.1. 确立“辅助为主、审查为要”的应用模式

生成式人工智能可应用于司法领域，但其定位必须基于“辅助而非取代”的原则。该项定位源于司法活动的复杂性和法律权威的不可替代性，也源于当前生成式人工智能的技术局限性。算法的决策逻辑受到数据质量和模型解释力的制约，可能对敏感案件产生偏倚或失真判断，故而人工智能在法律范围内的应用必须受到严格审查。司法裁判追求效率，更承载公正与可信度的使命，因此不能让机器决策成为裁判过程的唯一驱动力，应在技术赋能的背景下强化法官作为最终责任主体的功能。法官、检察官和律师作为司法裁判中的核心人员需对生成式人工智能的建议进行全面评估——审查其合理性、核验其合规性与证据的真实性，确保对任何虚假、偏倚内容的及时剔除[8]。

从技术视角而言，生成式人工智能的应用效果取决于司法数据库的完备性和真实性。司法数据是智能算法深度学习与生成能力的核心输入，若数据出现质量欠缺或来源割裂将直接影响案件推理的准确度与辅助决策的可信度。不同地域法院、不同部门、不同司法层级之间的信息孤岛现象，有可能导致数据共享不足或者输入样本断裂，大幅削弱人工智能综合信息处理能力。在此背景下需构建系统化且可追溯的司法数据共通机制，通过统一规范的采集、清理、更新及标准化操作确保数据的可靠性，针对性地解决法律知识更新、问题话语准确性以及法律知识储备问题，并探索实现生成式人工智能从通用到垂直的

专用化转变，经指令微调、指使学习、反馈学习、合成数据学习和低秩适配(LoRA)等方法，将目前大模型转变为法律领域大模型(见表 1)。

**Table 1.** Design logic and application link table of the auxiliary + review mode  
**表 1.** 辅助 + 审查模式的设计逻辑与应用环节表

关键环节	核心内容	司法人员角色	生成式人工智能角色
数据采集与清理	收集跨部门、跨区域数据库数据，清理错误、无效信息	制定数据采集规范；监督数据质量；参与模型设计	深度分析海量数据；自动筛查冗余信息
裁判建议生成	生成特定的文本、图形、音视频或模型，为法官提供“辅助手册”	审查建议的合法性与合理性；剔除问题数据	通过算法快速整合法律条文与往例案例，提供自主生成内容
辅助评判流程	法官结合生成式人工智能结果与案件具体事实独立裁定	独立评估机器建议；核验结果合规性；判定裁判权是否被过度干涉	生成可视化材料(如证据分析图表)；辅助类案比对
安全与透明	建立数据记录和可追溯机制，向公众公开部分裁判技术过程	监督透明记录的唯一性与准确性	保存算法记录；披露部分算法运行逻辑
裁判监控与反馈	法院定期审查算法运行状态，对裁判建议进行校验	终审意见负责人；反馈系统质量与不足	根据反馈优化算法模型，调整推荐逻辑

#### 4.2. 平衡算法与价值的互动，坚守“人机协同”

生成式人工智能依靠大数据和算法逻辑进行高效计算，能够在对证据材料的快速分析、类案检索、法条推送等方面大幅提升审判质效，但对机器决策的过分倚赖也隐藏法官价值判断空间被弱化的风险。司法不只是一种技术性的推理过程，更是一种深植伦理与情理的活动，因此需要在算法的演绎精度与人类司法者的综合判断力之间实现动态平衡。一方面，应当修复智能算法的内在缺陷，通过信息披露或对外部监管机构的安全评估，来确保算法数据的偏差、歧视与“黑箱”问题得到有效修正；另一方面，还要针对算法产生裁判建议时难以完全适应案件复杂性与人文需求的情况，保留足够的纠错和质疑程序。例如设置数据质量筛查机制、为偏差数据建立“清洗”流程，并鼓励对算法依赖度过高的情形进行审慎干预[9]；同时，对裁判系统内部人员和相关专业团队明确分工，避免技术设计被少数主体操纵，或因程序开发失误而造成不良后果。通过人机协同与价值筛选的有机结合，既能防止算法让渡裁判权带来的恣意风险，也能借助系统的计算力有效完成难点案件的事实认定和规则运用。

#### 4.3. 完善立法规制与司法伦理约束双轨并行

生成式人工智能融入司法裁判的过程应置于完备的程序和伦理约束体系之下。立法层面需在已有的《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》等基础上，进一步明确司法审判中人工智能应用的边界、程序与责任，规定相对独立的审查程序和监管环节，使相关算法或产品的合同方、设计方承担强制性的信息披露义务。技术规制应着眼于成果的输出，强调对算法研发与系统部署的过程干预，落实数据源头与算法模型的准入审查，构建全面覆盖“数据-算法-智能裁判结果”的风控方案。与此同时应加大在司法范围内的专业化人才培养，使审判人员能甄别深度合成及算法风险，最大程度减少认知障碍带来的不确定性。并同时兼顾分级监管理念，参考美国等国家分场景、分领域的做法，对不同类型的诉讼环节和技术应用予以差异化的规范与监督，以更灵活的方式平衡效能与安全、

创新与正义。

从司法伦理角度而言,应将隐私与个人信息保护贯穿于司法人工智能应用的全流程,以防止技术滥用或脱序而产生的社会不安。设计被侵害后及时救济和责任追究机制,避免“无人担责”或“多头推诿”现象。尤其要关注因算法信息泄露、数据歧视或偏差导致执法不公而侵害当事人合法权益的情况,将法官的独立思考与责任意识真正内化于裁判过程。在此基础上须落实对紧急处理程序与安全应急预案的明确规定,一旦智能系统在审判环节出现严重故障或异常时应当及时暂停,以防对案件当事人和司法公信力造成难以挽回的损害。

## 5. 结语

在推动司法效率和公平的同时如何妥善应对技术应用中的伦理和法律风险,将成为司法革新和保护公正的关键。研究通过分析技术赋能与司法革新的时代需求、风险挑战以及相应的解决策略提出构建一个“辅助为主、审查为要”的模式,为生成式人工智能在司法中的合理应用提供了理论依据和实践指导。

## 参考文献

- [1] 支雪婷. 生成式人工智能训练数据合理使用的适用问题探析[J]. 传播与版权, 2025(4): 107-110.
- [2] 郭金良. 生成式人工智能服务中“信息错误”的民事责任[J]. 政法论坛, 2025, 43(2): 25-35.
- [3] 张阿澜, 崔汪卫. 公正审判视野下人工智能审判的弊端与纾解之策[J]. 铜陵学院学报, 2025, 24(1): 71-75, 91.
- [4] 张凌寒. 深度合成治理的逻辑更新与体系迭代——ChatGPT等生成式人工智能治理的中国路径[J]. 法律科学(西北政法大学学报), 2023, 41(3): 38-51.
- [5] 陈吉栋. 人工智能法的理论体系与核心议题[J]. 东方法学, 2023(1): 62-78.
- [6] 周维栋. 论司法类案的效力层级及其冲突协调规则[J]. 中国法律评论, 2022, 47(5): 57-70.
- [7] 刘艳红. 人工智能的可解释性与AI的法律责任问题研究[J]. 法制与社会发展, 2022, 28(1): 78-91.
- [8] 张玉洁. 智能量刑算法的司法适用: 逻辑、难题与程序法回应[J]. 东方法学, 2021(3): 187-200.
- [9] 孙海波. 重新发现“同案”: 构建案件相似性的判断标准[J]. 中国法学, 2020(6): 262-281.