

人工智能辅助行政决策的法治风险及化解路径

贾霄燕

河北地质大学, 中国·河北 石家庄 050031

摘要: 随着数字政府建设的纵深推进, 人工智能技术已深度嵌入行政决策流程。算法权力的介入在优化裁量基准、压缩决策周期的同时, 也引发了行政主体资格虚化、正当程序消解、责任链条断裂及算法歧视等系统性法治风险。本文剖析人工智能辅助行政决策的运行机理与法律属性, 结合执法与公共服务场景检视其实践困境; 通过考察欧盟风险分级监管与美国柔性指引模式, 提炼契合我国体制的化解启示; 最终主张以风险分级为核心逻辑, 从明晰权责边界、重构差异化程序干预、健全责任分配与救济机制、强化全周期技术治理四个维度, 构建兼顾技术赋能与权利保障的协同化解路径, 为《行政程序法》制定与数字法治政府建设提供制度方案。

关键词: 人工智能; 行政决策; 法治风险; 风险分级

Legal Risks and Mitigation Pathways of AI-Assisted Administrative Decision-Making

Jia Xiaoyan

Hebei GEO University, China Hebei Shijiazhuang 050031

Abstract: With the advancement of digital government construction, AI technology has become deeply embedded in administrative decision-making. The intervention of algorithmic power, while optimizing discretion benchmarks and compressing decision-making cycles, has also triggered systemic rule-of-law risks, including the hollowing out of administrative subjects, the erosion of due process, the rupture of accountability chains, and algorithmic discrimination. This paper analyzes the operational mechanisms and legal attributes of AI-assisted administrative decision-making, examines its practical dilemmas in law enforcement and public service scenarios, evaluates the EU risk-based regulatory model and the US flexible guidance model, and proposes a risk-stratification-based resolution path covering responsibility allocation, procedural intervention, relief mechanisms, and full-cycle technical governance.

Keywords: Artificial intelligence; Administrative decision-making; Rule-of-law risks; Risk stratification

0 引言

人工智能进入行政领域, 最直接的变化发生在事实筛选和风险排序环节。过去由工作人员逐件查验的材料, 现在可以先由系统完成比对; 过去依赖经验判断的风险线索, 也可能由模型给出优先级。《法治政府建设实施纲要(2021—2025年)》要求推进政府履职数字化, 各地随之发展出智慧执法、信用画像和智能审批等实践。数字法治政府建设必须处理数字行政与法治行政的关系。^[1] 这些做法提高了行政效率, 却也使行政决定的形成过程更加复杂。算法输出如果只是内部参考, 风险相对有限; 一旦它成为处罚、给付或者许可的关键依据, 相对人就需要知道系统判断从何而来。算法系统嵌入公权力运行后, 技术判断会影响事实筛选和处理顺序。本文据此讨论人工智能辅助行政决策的法律属性, 并在应用场景和域外经验的基础上提出化解方案。

1 人工智能辅助行政决策的运行机理与法律属性

人工智能辅助行政决策的过程大致分为数据取得、模型处理和结果使用。行政机关将历史记录、申报材料 and 共享数据输入系统, 模型形成评分、预警或建议, 承办人员再决定是否采纳。数据取得阶段涉及多源异构数据的汇聚与清洗, 这一环节本身即蕴含价值选择——哪些数据被纳入、哪些被排除, 直接决定了后续模型判断的基线与盲区。模型处理阶段涵盖特征工程、算法训练与阈值设定, 其中特征选择体现着设计者对相关性 with 因果关系的理解, 阈值高低则反映了系统对假阳性与假阴性错误的不同容忍度。结果使用阶段表现为系统输出进入行政案卷并影响最终决定。这一过程改变行政判断形成路径, 尤其非现场执法、税务风控等高频事项中, 系统记录先于人工审查进入案卷, 形成了“技术前置过滤”效应——算法已完成对事实的初

次筛选,人工审查往往只在系统圈定的范围内进行,这在认知心理学上被称为“锚定效应”,即早期输入的信息会不当地限制后续判断的调整幅度。按人类介入程度,可分为人机协同和自动处理两类,我国目前以前者为主,但某些规则明确事项已接近自动化处理。法律评价应追溯数据来源、模型判断及人工复核是否真实发生,而非仅以“系统辅助”名义回避审查。

关于法律属性,学界有“算法主体说”“工具延伸说”“混合代理说”。^[2]将算法拟制为行政主体不符合我国行政法结构——算法无独立职权、意志和责任能力,无法成为国家赔偿责任的承担主体。完全视作为中性工具亦不足,因模型进入公权力后会改变事实发现和裁量顺序,其参数设定和特征权重已嵌入了政策偏好和判断倾向,并非纯粹的“中立计算”。较妥当的定位,是把算法理解为行政机关行使职权时使用的技术性判断装置。这一装置的深层意涵在于:装置本身不创造新的权力,但改变了权力的行使方式和使用条件,从而在事实上影响着裁量权的分配与运行。正如勘验、鉴定等传统行政辅助手段需要对其结论进行法律上的重述和评价,算法的输出也必须经过行政主体的“转译”才能成为法律意义上的判断依据——这种转译不是简单复述系统结论,而是要求承办人员将技术语言转换为法律语言,将概率判断重构为事实认定。行政机关可借助系统提高效率,但不能以此免除审查义务。

算法工具属性不降低法律控制必要。模型训练需选择数据,参数设定体现政策偏好,高权重指标将特定事实转化为高风险分值。例如,税务风控中系统可能将“短期内多次开具发票”赋以极高权重,但这一特征对批发零售行业恰是正常经营模式,若不加以审查,算法偏好便直接决定了稽查对象的选择。传统裁量要求说明事实、法律和比例考量,算法介入后这些判断部分隐于技术流程。行政机关需将模型输出转译为行政法上的事实认定和裁量说明,此转译过程应入卷,至少保留系统提示、人工判断与处理结果的对应关系,使后续审查能够还原技术参与的实际程度。

2 人工智能辅助行政决策的应用情形

智能系统已越过试点阶段,成为多领域行政活动的常规支撑工具。各部门使用算法目的不同——证据采集、线索筛查、资格审核或流程调度,功能差别带来风险差别,不宜笼统评价,需要根据不同的适用情形具体分析。

2.1 交通管理与非现场执法

电子警察和卡口记录车辆信息,推送违法图像。对闯

红灯等事实清晰的事项,系统记录常为主要证据。但图像识别不等于事实认定:雨雾、遮挡、标线磨损、施工改道等因素均可能影响识别准确性。实践中已出现因系统未能识别道路施工标识而误判的案例,也有因号牌污损导致处罚错误推送至他人名下的情形。若承办人员仅按系统提示批量处理,程序保障便严重滞后。交通场景中需明确系统适用范围,对标线模糊、设备遮挡等情形排除自动处理,保留异常筛查和异议复核通道,在决定书中说明证据来源。设备故障多发路段应设人工抽查比例,对夜间或极端天气条件下的图像采集设置更高的复核门槛。

2.2 智慧税务与风险筛查

税务系统用算法发现异常交易、确定稽查重点,综合发票、资金、社保、用能等多维数据生成风险评分。风险提示不直接构成处罚,但直接影响是否启动约谈或稽查程序。审查重点应放在风险提示转化为正式行政措施的过程——税务机关须说明触发风险的主要数据类型及其与法定稽查标准的关联。算法筛查应限定为线索来源,进入稽查后须由人员结合账册、合同等原始材料重新独立判断。对经常误判的行业,应建立年度分层验证机制,发现系统误差超过合理阈值时及时调整特征权重与评分模型,避免历史执法偏差通过训练数据代际传递。

2.3 社会保障与公共服务

给付行政中算法风险最为敏感——低保、医保、住房补贴面向弱势群体,错误停发直接。基本生存权益。系统通过公安、房管、银行等多源数据发现财产或资金变动后推送复核提示,但数据本身不能完全解释真实处境,如代持房产、临时性借款、共同居住人的非稳定收入等情形均可能被系统误读为“资产超标”。据报道,曾有低保人员因被冒名登记为企业“监事”而被系统自动停发补贴,经数月申诉才得以纠正。停发前应明确说明疑点事项,给予相对人补充材料和口头说明的充分机会,对不熟悉线上操作的申请人保留线下人工渠道。救助类算法应设置“人工复检安全阀”,凡涉及基本生活保障的自动停发决定,须经两名以上工作人员分别独立复核后方可生效。

2.4 审批监管场景

企业登记、资质许可、信用评价中,规则引擎完成条件比对并推送疑点;网络交易、食品安全、特种设备监测亦通过算法生成检查线索。问题在于规则库更新未必同步地方规范和行业标准的变化,跨部门共享数据的口径与时效也可能不一致。行政机关须回归法定条件独立判断,审批人员应在案卷中说明采纳或排除系统提示的具体理由,

不得将“系统预警”直接等同于“存在违法嫌疑”。

各场景表明,风险非单点出现而是系统性分布。法律控制应提前嵌入数据来源说明、复核记录和异议机制,将技术参与环节全面写入案卷,为后续审查提供可追溯的完整记录。

3 人工智能辅助行政决策的法治风险

算法介入后,传统行政法原则面临连续性压力。风险从数据取得阶段开始积累,并在运行、复核和救济中持续显现。

3.1 行政主体资格虚化

传统决定由法定机关作出,外部责任由该机关独立承担。算法介入后,数据提供方、算法开发者、系统运维单位均参与决定形成过程,行政机关却易将不利结果解释为“系统判断”,供应商则强调自身仅为技术工具,各方责任相互推诿,使相对人难以锁定有效的救济对象。更深层的问题在于,算法通过数据权重设置与分类阈值调整,实际上在资源配置和风险认定上拥有了实质话语权,这种技术权力进入处罚或给付程序后,便转化为相对人必须承受的法律后果。现行法律虽未细化此类情形下的归责顺位,但基本原则仍应坚持行政机关对外承担最终责任,供应商的缺陷通过合同追偿和民事赔偿渠道处理。

3.2 算法黑箱对正当程序的冲击

深度学习模型的参数规模动辄数以亿计,承办人员自身亦难以说明每项权重的具体影响,遑论向相对人作出清晰解释。说明理由义务要求行政机关使当事人知悉事实、依据和处理理由。^[3]算法将争议焦点推至数据来源和系统逻辑层面,传统听证程序若仅围绕最终结果展开则流于形式。^[4]可理解说明不要求公开源代码,但须说明主要数据类别、关键判断依据及人工复核情况,明确系统提示在最终决定中所处的地位。若系统输出为综合评分,行政机关至少应告知相对人哪些因素对结果起到了决定性作用,而非仅给出一个不可分解的风险分值。

3.3 算法偏见与公平性危机

偏见隐藏于训练数据和评价指标之中。历史执法集中于特定区域或群体,模型便将高检查频次误读为高风险特征;信用评价、资源分配亦受过往投诉记录、收入水平、区域统计数据的无形塑造。更为隐蔽的是“代理歧视”:算法虽不直接使用种族、性别等敏感特征,但可通过邮政编码、消费习惯、社交网络等代理变量间接产生近似的分类效果。模型将复杂的社会处境压缩为单一数值,必然忽视个案背景与比例原则的要求。当模型结论用于救助资格评

定、教育资源配置或信用限制时,个别误差便经年累月地沉淀为群体性的结构性不利。^[5]公平性治理要求行政机关在个案中保留充分的重新判断空间,并建立定期公平性审计制度,系统审查模型在不同社会群体间的误判率是否存在统计上显著的差异。

3.4 司法审查标准不明确

相对人仅能看到最终决定,无法接触模型版本、训练数据集和参数设定;行政机关常以信息安全或商业秘密为由拒绝披露,审查机关难以判断模型依据的合法性与可靠性。现行行政诉讼制度缺乏针对算法决策的专门证据规则,数据代表性测试与偏见检验尚无程序法上的依据。^[6]可建立有限证据开示制度,由行政机关提交系统版本号、复核记录和主要数据来源说明;涉及技术秘密的部分由专家辅助人或技术调查官在保密义务下审查。在证明责任分配上应区分两种情形:相对人仅对算法一般可靠性提出质疑时,由行政机关承担系统合规性的证明责任;相对人主张个案中算法输出存在具体误差时,由其提出初步证据后,由行政机关说明该案中系统输出与人工复核的关系。这种分层安排既避免过度加重相对人负担,又防止举证责任完全倒置带来的诉讼结构失衡。

4 域外经验及其对我国的启示

4.1 欧盟

欧盟以基本权利保护为中心,高度重视对自动化决策的事前控制。GDPR 第 22 条赋予数据主体在仅凭自动化处理作出重大影响决定时要求人工介入并表达意见的权利。《欧盟人工智能法案》采用风险分级思路,将公共服务分配、执法预测、社会保障等纳入高风险系统范畴,要求开展基本权利影响评估、全程保存操作记录、确保有效的人类监督,监管要求贯穿系统从市场投放前到投放后的全生命周期。该模式的优势在于边界清晰,将有限的监管资源精准配置于影响重大权益的系统;局限在于合规审计的文档负担较重。我国借鉴时宜区分层级:内部统计类系统适用备案抽查,涉及处罚、给付或资格认定的系统适用完整评估程序。

4.2 美国

美国未形成统一的联邦算法立法,主要依赖政策指引、技术标准与个案司法审查。2022 年《人工智能权利法案蓝图》提示公共部门关注安全性、防歧视与人工替代权;NIST《风险管理框架》要求持续风险识别、压力测试与第三方独立审计。法院依托《行政程序法》与正当程序条款,在个案中逐步累积对算法逻辑可质证性、结论不排

斥人类裁量的要求。纽约市 2021 年地方立法要求雇主对自动化就业决策工具进行偏见审计并公开结果,为公共部门算法治理提供了可参照的地方样本。该模式的灵活性较高,能适应技术快速迭代,但软法机制依赖机构自律,执行力度参差不齐,事后诉讼成本较高。

4.3 启示

欧盟模式侧重事前权利影响评估,美国模式侧重技术标准与事后司法校正。我国宜兼取二者之长:事前建立风险评估与备案制度,运行中强制要求操作日志与人工复核,争议发生后引入专家辅助审查与有限证据开示。算法治理应以风险分级为制度主轴,分级规则应明确写入《行政程序法》,使其成为可执行、可主张的法定程序权利。同时,我国可探索“框架性立法+技术标准+行业指引”的多元规范体系,在保持法律约束力的同时为技术发展预留弹性空间。

5 我国人工智能辅助行政决策的法治风险化解路径

5.1 明确责任分配

立法应确立行政机关最终负责原则,凡涉及人身自由、重大财产权益或基本生活保障的决定,一律不得完全交由系统自动作出。行政机关承担全面的审查、解释与纠错义务,供应商的技术缺陷通过合同责任、追偿或民事赔偿渠道处理。可引入过错推定规则:相对人证明决定存在明显异常或程序瑕疵后,行政机关应提交复核记录、系统版本号及数据来源说明;拒绝提交或无法作出合理解释的,承担举证不能的不利后果。政府采购合同中应明确约定数据质量标准、日志保存期限及故障应急响应要求,并将算法审计权列为合同必备条款。

5.2 构建风险分级程序干预

程序义务的强度应与风险等级相匹配。低风险应用(内部统计、政策模拟)适用备案加年度抽查,仅需保留系统使用记录备查;中风险应用(常规处罚、资质初审、信用评分、补贴发放)须在决定书中附具可理解说明,载明系统提示的主要内容、人工复核过程及采纳或排除的具体理由,同时为相对人提供便捷的复核申请通道;高风险应用(涉及人身自由、大额财产处罚、低保资格剥夺、出入境管理、重大环境健康事项)须在系统上线前完成合法性审查与伦理审查,运行中实施全流程留痕并定期出具独立审计报告,决定作出前原则上应举行听证并引入技术专家参与质证。风险等级应定期复核,系统功能扩展时应相应升级程序要求。算法影响评估应纳入法定前置程序,系统

评估数据来源合法性、样本代表性、使用目的是否超出授权范围以及是否存在可行的非算法替代方案。评估报告摘要可向社会公开,涉及国家秘密或商业秘密的部分作脱密处理。评估结论应作为系统上线和持续运行的前置性条件,未经评估或评估未通过的系统不得投入使用。上述程序规制的具体设计,可参考行政法上对自动化决策的程序要求^[7]。

5.3 健全救济机制

法院系统应建立技术调查官与专家辅助人制度,为涉算法案件的证据审查提供专业技术支撑。相对人有权申请第三方机构进行独立的算法审计,审计报告经质证后可作为诉讼证据使用。行政复议机关应设立专门通道,对涉算法争议案件快速受理,并依职权要求行政机关提交完整的复核记录与数据来源说明。对于系统性歧视或侵害不特定多数人权益的情形,应拓展行政公益诉讼的适用范围,由检察机关或符合条件的社会组织提起诉讼,要求纠正算法偏差并公开评估摘要。在证明责任分配上,相对人提出合理怀疑后,举证责任即转移至行政机关,由其证明系统具备可靠性、公平性与可解释性。救济程序中产生的专家鉴定费用可纳入诉讼成本范围,由败诉方承担,防止经济因素阻碍相对人获得有效的专业支持。

5.4 强化全周期技术治理

由国家数据局牵头建立全国统一的政府算法应用登记平台,实现分类登记、版本管理与动态更新。任何算法系统未经安全评估、伦理审查与合规测试,不得投入影响公民权益的业务运行。高风险系统须接受第三方独立审计机构定期检查,重点审查样本偏差率、误判率、模型鲁棒性及复核记录完整性。伦理审查委员会应由法学、计算机科学、公共管理等领域专家及公众代表共同组成,系统评估数据使用边界与潜在的歧视性影响。建立动态退出机制:凡连续出现系统性偏差、引发重大社会争议或无法通过合规审查的系统,应暂停使用并限期整改;技术架构严重落后且无法满足现行标准者,依据日落条款有序退出。系统采购环节应将算法透明度与可审计性列为关键技术指标写入招标文件,防止行政机关在引入技术时即丧失后续监管能力。对已投入运行的存量系统,应设定合理的过渡期完成合规改造,避免新旧制度之间的监管断层。

6 结语

AI 辅助行政决策已是不可逆转的现实。它在提升行政效率的同时,也可能削弱主体责任、压缩程序参与空间、增加权利救济难度。算法只能辅助处理事实与线索,绝不

能替代法定行政机关完成价值判断与利益权衡。技术介入越深,行政机关就愈应承担与之相匹配的解释、复核与纠错义务。我国应以风险分级为制度主线,将不同强度的程序义务精准配置于不同风险等级的应用场景之中。未来《行政程序法》应设立专门条款,明确行政机关最终负责原则、高风险系统必须保留人工决定权,并赋予相对人获得可理解说明、申请复核和提出异议的完整权利束。条款设计应至少涵盖以下内容:规范的适用范围界定、风险分类标准及对应的差异化程序义务、行政机关的数据质量保障与日志保存义务、相对人的知情权与异议权及获得人工复核的权利、第三方独立审计与专家辅助的启动条件。责任规则厘清行政机关与技术提供方之间的关系,救济规则引入专家辅助与有限证据开示。与此同时,应密切关注生成式人工智能等新兴技术的演进趋势,为制度框架预留充分的适应性调整空间。唯有将算法的效率优势切实转化为可解释、可复核、可问责的行政能力,AI 辅助决策才能真正服务于数字法治政府建设,在技术赋能与权利保障之间建构起均衡而稳固的制度支点。

参考文献:

[1] 王锡锌. 数字法治政府建设的底层逻辑及展开路径

[J]. 数字法治, 2023(02): 7-11.

[2] 张凌寒. 算法权力的兴起、异化及法律规制[J]. 法商研究, 2019, 36(04): 63-75.

[3] 展鹏贺. 论数字行政正当程序构造[J]. 中国法学, 2025(05): 104-122.

[4] 唐安然. 自动化行政中正当程序原则的重构[J]. 比较法研究, 2025(05): 177-191.

[5] 蒋红珍. 数字政府建设面临的法治议题: 以判例为观察视角[J]. 中国法律评论, 2025(06): 36-52.

[6] 房娇娇. 生成式人工智能辅助行政决策的算法隐患及其治理路径[J]. 湖湘论坛, 2024, 37(01): 99-111.

[7] 吴锐, 张若涵. 自动化行政中对人工智能算法决策的程序规制[J]. 成都理工大学学报(社会科学版), 2022, 30(05): 58-67.

基金项目: 该论文是河北地质大学区域制度研究中心项目的结项成果。

作者简介: 贾霄燕(1978.02-), 女, 汉族, 河北邯郸人, 硕士, 副教授, 研究方向: 行政管理和行政法治。