Open Journal of Legal Science 法学, 2026, 14(1), 167-172
Published Online January 2026 in Hans. https://www.hanspub.org/journal/ojls
https://doi.org/10.12677/ojls.2026.141022

Hans汉斯

法律应对个性化推荐算法隐私侵蚀的路径转向

钟东海

宁波大学马克思主义学院，浙江 宁波

收稿日期：2025年12月12日；录用日期：2025年12月29日；发布日期：2026年1月14日

摘　要

面对个性化推荐算法对隐私的系统性、持续性侵蚀，传统以“用户知情同意”为核心的前端、静态法律框架已然失效。法律应对的路径必须实现根本性转向：从事后救济、形式合规，转向以“算法问责”为核心，通过强制性的算法审计、持续的风险评估与透明的信息披露，构建一个动态、过程化、以规制算法系统本身为重心的新型治理范式。

关键词

个性化算法推荐，隐私侵蚀，法律转向

A Shift in Legal Approaches to Privacy Erosion by Personalized Recommendation Algorithms

Donghai Zhong

School of Marxism, Ningbo University, Ningbo Zhejiang

Received: December 12, 2025; accepted: December 29, 2025; published: January 14, 2026

Abstract

Faced with the systematic and persistent erosion of privacy by personalized recommendation algorithms, the traditional front-end, static legal framework centered on “user-informed consent” has become ineffective. The legal response must undergo a fundamental shift: from post-remedy and formal compliance to a new governance paradigm centered on “algorithmic accountability”, which involves mandatory algorithmic audits, continuous risk assessments, and transparent information disclosure, aiming to build a dynamic, procedural governance paradigm that focuses on regulating the algorithm system itself.

文章引用: 钟东海. 法律应对个性化推荐算法隐私侵蚀的路径转向[J]. 法学, 2026, 14(1): 167-172.
DOI: 10.12677/ojls.2026.141022

Keywords

Personalized Recommendation Algorithm, Privacy Erosion, Legal Shift

1. 引言

在数字经济浪潮与技术权力的深度重构下，个性化推荐算法已成为信息分发的核心引擎，其在提升效率与体验的同时，也深刻重塑了隐私保护的逻辑与边界。学界日益认识到，算法可能将用户置于数字“圆形监狱”之中，使其隐私空间趋于透明、控制力被消解。近年来，围绕数字经济治理、算法推荐机制与个人信息保护的研究涌现，对算法“黑箱”及其侵害责任的探讨，以及对“知情同意”原则有效性的反思，已成为学术焦点。然而，现有研究多集中于现象揭示或局部规制，对隐私侵蚀的系统性机理剖析，以及超越“知情同意”、迈向全过程治理的范式转型论证，尚缺乏贯通性的理论整合与制度建构。

本文致力于深入揭示个性化推荐算法隐私侵蚀的内在机理，系统论证传统“知情同意”框架的结构性失灵，并构建以算法审计为核心的规制新范式，试图为理解算法时代的隐私保护困境提供一个机制性的解释框架，并为构建能够穿透技术黑箱、制衡算法权力的法律制度贡献理论依据与路径参考。

2. 个性化推荐算法隐私侵蚀的机理

个性化推荐算法将为信息监控提供便利，甚至成为大规模远程监控的工具，让用户深陷算法、数据、算力等数字力量构架的数字“圆形监狱”，个人隐私空间逐渐趋于透明，用户对个人隐私的控制力被消解[1]。个性化推荐算法对隐私的侵蚀，是一个“从微观行为到宏观画像，从静态知情到动态操纵，从个体侵害到系统性风险”的深刻过程。它不再仅仅是关于“信息被谁知道”的秘密问题，而更关乎“个体如何被评估、预测和塑造”的自主性与尊严问题。这种机理上的根本性变化，宣告了仅仅依靠赋予用户前端控制权的法律路径的失效。因此，法律必须将目光从用户同意的“那个瞬间”，转向算法系统内部运行的“整个过程”，寻求一种能够穿透技术黑箱、规制系统性风险的新范式。

2.1. 数据收集的无感化和全景化

传统隐私侵犯多涉及对明确声明的敏感信息的盗取或泄露。而推荐算法的隐私侵蚀始于一种更隐蔽、更广泛的数据收集。其一，行为数据成为“新石油”：每一次点击、滑动、暂停、快进，甚至鼠标移动轨迹，都被转化为可分析的行为信号。这些数据本身看似非敏感信息，但却是构建数字人格的原始材料。其二，情境数据的无孔不入：算法同步收集设备信息、网络环境、地理位置等，将用户行为置于特定的时空情境中加以解读，极大地增强了数据的可关联性。其三，“无感收集”成为常态：用户为获得即时、便捷的服务，往往在无明确感知的情况下持续“喂养”算法。这种收集是持续的、背景式的，而非一次性的、需要用户主动提交的，使得传统的“收集时点告知”变得意义寥寥。

2.2. 数据处理与推断的“黑箱化”与“关联化”

“黑箱化”指数据处理过程的不透明性，即算法决策机制复杂且难以解释，决策逻辑对用户或监管者隐藏。机器学习模型通过海量数据训练后，其输入与输出之间的映射关系可能无法用简单规则描述，

导致“算法黑箱”[2]。这是算法隐私侵蚀的核心环节，也是法律最难规制的部分。算法通过协同过滤、嵌入表示等机器学习技术，在海量用户数据中寻找潜在模式。例如，通过分析 A 用户与 B 用户在成千上万个物品上的相似行为，即使 A 从未透露其政治倾向，算法也能因其与 B 的高度相似性，推断出 A 的政治偏好，并将其标签化。这是对隐私的深层突破。通过点赞、浏览记录等非敏感数据，算法可以高精度推断出用户的种族、性取向、宗教信仰、智力水平甚至心理健康状况等核心敏感信息。这种推断性信息的生成，完全绕过了法律对“敏感个人信息”需获取“单独同意”的规定，因为它并非“收集”而得，而是“计算”而出。用户画像不是静态档案，而是一个实时更新的动态模型。算法会根据用户的最新反馈即时调整策略，形成一种“刺激–反馈–优化”的适应性循环。这意味着，对用户的了解和控制是一个不断深化的过程。

2.3. 输出影响的“操纵性”与“系统性”

隐私侵蚀的最终危害，体现在算法输出对用户认知和选择的塑造上。算法以“相关性”和“参与度”为最高目标，倾向于强化用户既有偏好，形成“信息茧房”或“过滤气泡”。这不仅限制了用户的视野，更使其在无形中丧失了接触多元信息、自主形成判断的机会，构成了对思想自由和自主决策权的侵蚀。更前沿的推荐系统致力于“塑造”而非仅仅“满足”用户偏好。通过精心控制信息暴露的序列和频率，算法可以潜移默化地改变用户的消费习惯、政治立场甚至情感态度。这种“助推”或“操纵”是高度个性化的，且通常以提升平台商业指标为目的，使隐私侵害与商业利益深度捆绑。如果训练数据存在偏见，或优化目标本身带有倾向性，算法会将历史上的歧视模式自动化、规模化。例如，在招聘、信贷或内容推荐中，算法可能系统性地边缘化特定群体。这种侵害不再是针对个体的、偶发的，而是嵌入系统逻辑的、结构性的。

3. 传统“知情同意”框架的失灵与局限

在个性化推荐算法所构建的数字环境中，“知情同意”框架遭遇了结构性的，而非技术性的失灵。问题的根源不在于告知不够清晰或同意不够频繁，而在于这套框架所依赖的个人主义、静态化、契约式的规制哲学，与算法系统的系统性、动态化、权力结构化的现实从根本上不匹配。

3.1. 知情同意的法理基础及其在算法时代的前提崩塌

“知情同意作为个人信息保护的神经中枢，系协调《民法典》中关于个人信息保护的规范与《个人信息保护法》之间关系的核心节点，亦是破解个人信息治理难题的一大关键基点。”[3]知情同意具有双重法理意义，亦是个人自决权的核心体现。这源自医疗法和数据保护法。其理想模型是，一个充分知情、理性且自主的个体，在自由意志下对特定数据处理行为做出授权。它预设了数据主体与处理者之间一种基于信息的、相对平等的对话关系。二是风险分配与责任划分的工具。在法律实践中，有效的“同意”构成了数据处理合法性的关键基石，也是处理者转移部分法律风险、证明其行为正当性的核心证据。

“知情同意”的有效性建立在信息对称的可能性、真实选择权和处理行为的确定性与有限性三个隐含前提之上，而算法环境使这些前提悉数崩塌。算法系统的复杂性使其决策逻辑远超普通人乃至专家的直观理解。平台提供的隐私政策，用概括性、技术性语言描述的是一个“可能性的宇宙”，而非具体的处理行为。用户面临的不是信息不足，而是信息超载与认知过载，真正的“知情”在事实上已不可能。在“要隐私”与“要服务”之间，用户面临的是实质上的“捆绑交易”。拒绝个性化推荐往往意味着服务功能残缺、体验急剧下降，甚至无法使用核心功能。这种权力结构的极度不对等，使得“同意”不再是自由选择，而是一种为获取数字生活必需品所被迫支付的“数据对价”。用户同意的是一份模糊的“授权书”，

而算法执行的是持续、动态、探索性的数据处理。初始同意的“目的”被无限拉伸，后续的数据推断、画像更新、跨场景应用等关键处理环节，均未在同意时点被具体界定。用户同意的是一个开放的、不断扩张的过程，而非一个封闭的、确定的行为。

3.2. “知情同意”的实践异化：从保护工具到合规仪式

平台设计的核心逻辑已发生根本性转向：其重点不再止于保障用户对条款的真实理解，而是聚焦于如何以最高效的方式获取符合法律形式要件的“同意”行为。这一演变实质上将“知情同意”机制从实质性的用户权利保障工具，转化为一种追求合规效率的流程优化对象，使其在实践中趋于“点击即合规”的形式化操作。在此模式下，用户的点击行为被转化为平台获取形式合法性的依据，从而使其得以在后续可能产生的隐私争议中，凭借该“自主授权”将相关风险归责于用户。由此，同意机制的功能发生异化：从原本保护用户权益的盾牌，转变为强化平台免责能力的盔甲。

面对持续涌现、内容繁杂且难以理解的隐私提示界面，用户的决策状态亦呈现显著变迁——从初期的审慎关注，逐渐演变为普遍的决策疲劳、倦怠与机械性同意。这实质上是一种理性支配下的行为放弃：当用户意识到自身无法在有限认知与时间内作出实质性判断时，便倾向于采取策略性回避，通过快速同意以继续后续操作。在此情境下，同意作为个体自主决定与意思自治的核心内涵已被严重掏空，其制度初衷与实践效果之间出现深刻断裂。

3.3. 法律修补努力的局限性与内在悖论

立法者与监管者已意识到上述问题，并尝试进行修补，但这些努力大多陷入困境。以分层提示、图标化、短视频说明的方法为例。这种试图使信息更友好。然而，它们只能简化对“收集什么”的描述，却完全无法解释“将如何算法化处理及产生何种影响”这一核心。这如同向病人详细描述手术刀的形状，却完全不解释手术将如何切割以及可能的风险。局限性在于触及了人类认知与复杂系统之间不可逾越的鸿沟。简化不等于可理解，尤其是当被简化的对象本身是一个不可简化的黑箱时。

修补性努力如同在泰坦尼克号上“选座位”，它们或许能缓解局部的不适，却无法改变船体正在下沉的命运。法律若继续将“知情同意”作为应对算法隐私侵蚀的首要乃至核心工具，无异于刻舟求剑。因此，我们必须承认这一范式的局限性，并积极探索将规制重心从用户端的形式授权，转向算法系统本身的实质问责的全新路径。

4. 路径转向的核心：从“知情同意”到“算法审计”

所谓算法审计，是指审计主体对被审计者所使用的算法模型、数据、算法研发等技术活动的风险、合规进行审核、评估，以监督算法正当使用的活动。“世界各国立法也多有提及算法审计，如美国纽约州立法明文规定雇主在招聘过程中使用算法时，必须进行年度算法歧视审计，否则不可在招聘过程中使用算法进行简历评估。”[4]

4.1. “审计”必须取代“同意”的理论基础

在风险规制理论中，其逻辑出发点有三类：“其一，将风险管理视为实现更好合规的手段。其二，为了确定哪一类风险需要规制，应综合考虑技术对健康、安全和环境的影响。其三，基于监管成本，对规制行为进行优先排序。”[5]算法推荐是一种具有持续高风险的数据处理活动，法律应将规制重点从用户端转移到数据处理端，要求其对算法系统的风险负持续的管理责任。借鉴产品责任与环境法中的思想，算法作为数字时代具有重大社会影响力的“产品”，其生产者与运营者必须为其可能造成的系统性社会危害承担责任。这种责任不能通过获取用户同意而完全转移或豁免。传统隐私法聚焦于对已发生个体侵

害的事后救济。而个性化推荐算法产生的是系统性、持续性且常发生于损害发生前的风险。风险规制理论要求法律采取预防性、前瞻性的姿态，对风险的制造者施加持续的风险管理义务。

借鉴产品责任思想，算法作为数字时代具有重大社会影响力的“产品”，其生产者与运营者必须为其可能造成的系统性社会危害承担责任。这种责任不能通过获取用户同意而完全转移或豁免。“信用治理动态调整使得信用治理机制具有规范互动的不确定性。需要树立规范技术运行、激励技术创新的共识，运用算法规制结合产品责任的治理逻辑，在‘融通并行’理念引导下，形成新型信用评分算法治理方案。”[6]

4.2. 新路径的基石：算法审计制的三大支柱

算法审计并非单一工具，而是一个由三大支柱构成的、相互支撑的治理框架。这三者共同作用，旨在穿透“技术黑箱”，实现持续的过程性监督。支柱一是有意义的透明度：作为信息披露义务的履行。它指的是在保护商业秘密与知识产权的前提下，算法运营者负有向监管部门、利益相关方及社会公众提供足以理解系统功能、逻辑与潜在影响的解释性信息的责任。其核心在于实现功能可理解性与行为可预期性，而非公开源代码等核心机密。通常包括公开算法的基本目的、核心输入数据类型、主要技术架构、训练数据的基本统计特征；说明算法设计时考虑的主要利益相关者、已识别的潜在风险及所采取的缓解措施；在自动化决策对个人产生重大影响时，提供清晰、及时地通知并指明有效的申诉渠道。该支柱直接对应算法运营者的法定信息披露义务。它是履行“告知–解释”责任的核心环节，构成了后续风险沟通、社会监督与审计验证的基础。在法律上，违反透明度要求可能直接触发行政监管责任，并构成民事领域侵犯知情权或认定存在过程性过错的证据。

支柱二是全周期的风险评估：作为过程性预防义务的体现。它是一种强制性的、文件化的管理流程，要求算法运营者在设计、开发、部署、运行及退役等关键节点，持续性地识别、分析、评估和缓解算法可能引发的伦理、法律及社会风险。涵盖设计阶段对算法目标合规性、数据来源合法性、潜在偏见及对基本权利影响的预评估；部署前在代表性测试环境中进行的公平性、准确性、安全性及鲁棒性验证；运行阶段建立的持续性监控指标与定期复审机制，以探测实际运行中产生的非预期歧视后果或其他社会危害。该支柱确立了算法运营者的勤勉尽责或过程性预防义务。它要求运营者建立并维护一套内部治理程序，以证明其已采取合理措施避免可预见的损害。风险评估所产生的文档记录，是证明其已履行该注意义务的关键证据，直接影响行政合规认定与民事责任中过错要件的判定。

支柱三是独立可验证的审计：作为监督验证与问责保障。它是指由独立于算法运营者的合格第三方机构或组织内部独立部门，依据既定标准，对透明度声明的真实性、风险评估过程的充分性及缓解措施的有效性进行系统检查、测试与评估的监督活动。包括对透明度报告内容的核实、对风险评估方法论与执行情况的审查、对算法实际输出进行技术测试以检测是否存在隐含偏见或违规行为，并最终形成公开或向监管机构提交的审计报告。该支柱是前两大支柱的校验机制与执行保障。它将运营者的自我声明转化为可被外部客观评价的对象，审计报告从而成为具有公信力的专业证据。在法律责任体系中，审计结论可直接作为行政机关采取执法行动的依据，或在司法诉讼中作为专家证据，用于认定运营者是否切实履行了其法定义务，从而为法律责任的最终落实提供技术支撑与事实基础。

4.3. 审计制度框架的运作逻辑：一个动态闭环系统验证

三大支柱环环相扣，共同构建起从“风险识别”到“义务设定”再到“责任承担”的完整法律逻辑链条：首先通过风险评估将技术性危险具体化为可管理的法律风险清单；继而基于识别出的风险，为运营者设定透明度解释、风险预防及接受审计监督的行为义务；最终在损害发生时，借助审计验证结论逆

向审查义务履行情况，为过错认定、因果关系判断及责任追究提供坚实依据。这一“治理–责任”一体化框架，推动算法问责从单纯关注损害结果转向同时强调对风险管理过程的规范与监督，从而建立起一个兼具前瞻性、过程性与可执行性的法律治理体系。

审计独立验证透明度声明的真实性与风险评估的充分性，并发现未被评估的风险。审计结果反馈给运营者，要求其整改；反馈给监管机构，作为执法依据；在脱敏后向社会公开，形成市场与社会监督压力。循环往复：促使运营者持续改进其透明度报告与风险评估实践，形成一个“披露–评估–审计–整改–再披露”的持续改进循环。“算法审计制”的提出，标志着法律应对策略完成了从“规范用户行为”到“规制技术系统”，从“追求形式合规”到“追求实质安全”，从“事后侵权救济”到“全周期风险治理”的根本性范式转型。

它不再依赖处于弱势的用户去理解并约束强大的算法，而是直指权力的核心：算法的运营者，通过一套结构化、程序化、可验证的制度安排，迫使其将隐私与基本权利保护内化为系统设计与运营的必然要求。

参考文献

[1] 谢永江, 杨永兴, 刘涛. 个性化推荐算法的法律风险规制[J]. 北京科技大学学报(社会科学版), 2024, 40(1): 77-85.

[2] 王莹. 算法侵害责任框架刍议[J]. 中国法学, 2022(3): 165-184.

[3] 马新彦, 黄舜. 论知情同意在个人信息保护规范中的属性[J]. 吉林大学社会科学学报, 2024, 64(5): 85-98+237.

[4] 张欣, 宋雨鑫. 算法审计的制度逻辑和本土化构建[J]. 郑州大学学报(哲学社会科学版), 2022, 55(6): 33-42.

[5] 曾雄, 梁正, 张辉. 中国人工智能风险治理体系构建与基于风险规制模式的理论阐述——以生成式人工智能为例[J]. 国际经济评论, 2025(4): 131-152+7.

[6] 杨帆. 信用评分算法治理: 算法规制与产品责任的融通[J]. 电子政务, 2022(11): 28-39.