

doi:10.16060/j.cnki.issn2095-8072.2026.04.006

算法权力的多维渗透与调适^{*}

须 露

(西南政法大学人工智能法学院, 重庆 401120)

摘 要: 在数字经济时代, 算法已成为塑造社会权力结构的工具, 它以技术理性的表象掩盖内在的价值偏向, 通过外化为社会规则、全景敞视的监控、自主决策等形式, 形成一种无形却系统的控制。然而, 数字主体并非完全顺从, 而是采取主动抵抗、被动防御及组织化抗议等反抗策略, 形成技术与用户的动态缠斗局面。这一过程暴露出社会规则与算法规则核心架构、用户个人信息保护与算法全景敞视主义、用户自主选择权与算法情境引导之间的深层矛盾。为应对算法权力治理挑战, 需理解算法内部核心运行机制, 以优化算法设计为基础, 对其内在运行机制进行源头治理, 弱化算法规训, 并以限定监视场景与数据采集范围、设置数字追踪提示为核心约束算法监视, 最后以内外部分散权治为手段弱化算法自主决策权, 最终达到算法权力规则的适配、用户基础权利的保障以及平台经济健康有序发展。

关键词: 算法权力; 核心架构; 算法监视; 数字缠斗; 算法规制

中图分类号: D912.29/F49

文献标识码: A

文章编号: 2095—8072(2026)04—0085—12

一、引言

数字技术的快速迭代与平台经济的深度渗透重塑了社会生产与生活的基本范式。算法作为这一变革的核心驱动力, 已从单纯的技术工具演变为新型权力载体, 凭借其无形性、渗透性和系统性特征, 重构着平台经济生态系统中的主体关系格局。它通过多数据的占有、处理与结果输出, 演化为资源、商品、财产、中介乃至社会建构力量 (Braman, 2006), 并逐渐演变为行为规范, 影响并控制着个体的行为, 甚至取代公权力, 作出具有法律效力的算法决策 (张凌寒, 2023)。随着大数据和人工智能算法服务的日益扩大, 算法权力正以一种悄然又迅猛的方式渗透到日常生活和社会运转之中。

当前针对算法权力的研究主要集中于三个维度。一是, 剖析算法权力的技术特征。有学者提出基于算法权力的机器、架构、嵌入三大优势可以归纳出算法权力的跨越性与隔离性两大特征。在这两大特征的驱动之下, 算法借由技术面纱逃逸现有的法律规制, 无形中又进一步巩固了算法权力 (张凌寒, 2021)。二是, 从本质论的视角揭示算法权力的控制本象, 即在计算机软件或程序中, 算法通常呈现为由逻辑条件和控制结构所构成的“逻辑+控制”的计算方案。以数字化的全景敞视监控与无形的算法规则, 形成技术与人类行为之间的控制样貌 (段伟文, 2018)。三是, 针对算法权力中的无序控制, 回归治理本质, 探寻算法权力的规制困境与路径。其中, 由算法权力不当控制引发的算法黑箱、算法歧视、个人隐私及数据安全等问题, 学者们尝试从注重算法设计的伦理性 (郑智航, 2021)、算法运行过程的可追责性 (肖红军, 2022) 以及增强算法可解释性 (袁曾, 2025) 等角度出发, 期冀达成对于算法权力不当控制的法律规制。

^{*} 基金项目: 本文受2024年国家社会科学基金项目“民事司法中的‘官批民调’研究 (1875—1926)” (项目编号: 24BFX161) 和重庆市研究生科研创新项目“数字‘幻影’捕手: 生成式人工智能幻觉识别标准与治理路径研究” (项目编号: CYB25191) 的共同资助。

但未及类型化的研究使得算法权力的窥见与治理停留在散落的边缘,既未能考虑到不同主体的控制方式差异,也未能认识到权力控制的动态性和互动性,从而盲目聚焦于形成以法律法规为基础、具有普适性的集中式治理模式上,强调全方位的规制,限制了算法的合理发展,实际上也不具备任何可行性。例如,算法权力对于数字生产者、数字劳工以及普通用户权力的无形侵蚀,从实践上逐渐引发各主体不同的抗争策略与互动博弈,最终形成用户与算法“越反抗越发展”的数字缠斗局面。

本文采用理论建构与实践分析相结合的研究方法,从算法权力的理论解构出发,系统分析其在平台经济中的多维渗透机制,深入揭示算法权力控制下各主体的抗争实践与动态博弈,进而创新性地提出算法权力与用户权利矛盾的调适框架。技术上,通过厘清算法权力控制的运行机制,提出了以优化算法设计为基础的源头治理方案,从而弱化算法对用户的刻意规训;理论上,通过整合理论批判与中国治理实践,构建了以“限定+提示”为核心的算法权力约束体系;程序上,倡导算法决策的分布式治理方案分散算法的自主决策权,为用户权利的保障提供可行性参考。

二、算法权力多维渗透的现实图景

人工智能、大数据、云计算等技术正在取代传统社会的生产方式和管理模式,对个人、平台企业、社会、国家均产生了重大影响。算法作为技术的底层逻辑,将其内在的算法规则外化为使用者的社会规则,同时凭借其突破时空限制的技术优势,对用户行为进行全方位监控,并在此基础上构建起基于用户个体的数据库,对用户的选择权进行隐性的干预与引导。这种持续的规训、控制与干预,使算法逐步超越单纯技术工具的属性,演变成为一种新型的权力载体,渗透于数字社会的各个角落。

(一) 算法规则从技术嵌入到社会规训

算法权力渗透的第一个维度表现为将算法规则外化为数字主体的社会规则。这一过程的本质是将特定经济或治理目标隐形转化为算法编码的逻辑,从平台内部映射于社会现实的外部行为准则。在技术渗透的初期阶段,算法作为一种技术工具,依托其高效率、低成本与高利润等优势被投入使用,但是随着智能化进程加快,算法使用的领域逐步外扩至金融、人力、司法辅助等关键领域,而物理世界运行的基本规则也随之内嵌于算法系统中,并在算法的加持下自动化、规模化地执行决策与资源分配,实质性地替代了传统的人工裁量和制度规范。此时,莱斯格(2018)所描述的“代码即法律”逐渐显现,算法的逻辑演变成为一种新型的、高效的、却不透明的社会规则。而这一由算法权力所导致的“规则外化”的首要后果是权力的“黑箱”,技术上的不可控使得算法与人的行为分离,进一步使人的行为与责任分离(吴椒军和郭婉儿,2021),传统法律中的“行为—责任”链条被算法权力规训所割裂,从而使传统的法律规制无法很好地作用于算法系统,算法摆脱被人所控制的工具属性而产生权力趋势(张凌寒,2021)。算法训练背后所依靠的海量数据,不可避免地放大乃至改变社会评价系统,个体为了在算法系统中生存或者获益,会不断追求高分、高曝光率,并据此进行自我审查与行为调适,即算法权力通过外化的规则渗透并内化为个体的主动遵从。此时,社会评价不再是传统的基于诚信、良知等基础上的道德共识,而是依赖于网络中所反映出的基础数据。例如,平台商家评分的高低会直接影响到消费者的行为选择,外卖骑手是否准时、是否好评、是否遵循算法的派单路径等也会影响骑手的等级。内容创作者必须迎合推荐算法的偏好,否则无法

获得系统的流量推送。于是，个体在无意识中将算法的隐性要求转化为日常实践，进一步削弱了对权力运作的感知与抵抗能力。这种规训不仅作用于行为表现层面，更深入至用户的认知判断与行为选择逻辑，使得所谓自主决策往往只是对算法预设路径的被动响应。与被动遵从不同，这种对个体用户的内在规训所对应的是高效能的算法技术，网络算法中重要的不再是签字或数字，而是代码，代码即数字主体行为的密码（蓝江，2021）。

（二）算法权力从效率至上到全景敞视的监控

算法权力渗透的第二个维度体现在超越时空限制对用户进行的“全景敞视”（panopticism）的监控（福柯，2019）。算法的超时空性正在并将持续重塑人类经由物理世界建构起来的关于时空的基本感知与体验。将网络空间按照纵向可以分为数据与存储分析、物理层、传输协议/域名系统、内容层、应用程序、操作系统、硬件终端/传感器、用户八个层面。每一个层面都可以通过“账户—数据—评分”机制加以改造，使实际控制力和服务效率不断提升。纵向分层本身拉伸了网络空间的维度，从日常可接触的终端到不可见的云端，由此创造了新型权力关系，即算法权力（胡凌，2024）。在时间维度上，网络实时性的追求极度地压缩了社会反应周期，使社会处于一种高速运转状态之中，同时，基于大数据的相关性预测，使得同一时间不同群体都会受相关性的涵摄而接收到同一信息，其时空的交错性也由此体现。除此之外，算法还利用自身的计算优势，通过地理位置、物流网络 and 平台调度，将物理空间数据化和可计算化，并在效率至上主义的加持下对物理空间进行重新规划与优化。而这些看似及时高效的普惠背后，藏着的是“核心—边缘”与“高效率—低性能”的时空鸿沟与价值对立的另一面，以时空跨域为优势的算法权力在跨越中造就新的不平等。信息来源的多渠道、多方位、多变化等特点也使得信息的接收方陷入“真假难辨”的认知牢笼，从而导致来自网络信息的信任瓦解消散。每个人被推送的算法时间（如不同的信息流）和算法空间（如不同的推荐地点）都被高度个性化后，就导致了共同的、同步的公共体验（如同时看一份报纸、一个电视节目）的消失，从而使社会共识难以形成。例如，每一次推送、每一次弹窗、每一次点击其实都是算法根据用户画像为其量身定做的视觉表演。在网络世界中，主导用户的不再是“他想看到什么”，而是“他想让你看到什么”，这也是为何相同的搜索词用户所搜索出的内容呈现巨大反差的原因所在。同时，算法借助无处不在的传感器和智能设备，实现了对个体时间和空间（线上与线下）的持续追踪，为权力的渗透提供了全天候的监视。在此过程中，算法系统会持续进行数据收集与评估，形成一种无形且有效的监管机制。用户不知自己何时被监视，但算法系统却会根据用户的选择偏好不断调整算法，形成一种数字化的“全景敞视”的监控（福柯，2019）。

（三）算法权力场景从自主决策到认知模塑

算法权力渗透的第三个维度是通过场景设置来实现其最隐蔽、最深刻的渗透。此种权力不再以直接命令或强制规则的形式出现，而是通过设计用户行动的网络环境来施加影响，其核心机制是通过信息茧房构建与偏好锻造实现控制。算法平台通过精心的界面设计、默认选项、信息排序和助推策略，巧妙地引导用户在看似自由的情况下做出符合平台利益的选择。实际上，系统构建了一个用户认知的行为引导机制，不知不觉地训练着用户的认知范围、偏好结构与行为模式，为每一个用户构建了高度定制化的情境现实。在各式各样的场景里，通过使用者的活动产生了大量的数据，这些数据被储存在对应的账户之中，因此，大量的场景都有了“账户—数据—评分”这样的微观权力结构，这意味着网络空间具有一定的同构性和趋同性。这也说明了在场景维度之下，网络空间未必是

天然的自由世界，而是由无数个独立架构相互连结构成的世界（胡凌，2024）。在一个个的场景设置中，个体所能看到的信息、所能接触的商品乃至所能互动的对象，均是算法设置和预演的结果（高进和黄种兴，2025）。这种控制的独特之处就在于既能满足用户的表层需求，又能引导其产生新的需求；既尊重用户的选择权，又影响其未来选择的方向，进而形成一种近乎完美的控制机制，使得抵抗变得困难，甚至使抵抗的必要性都难以被理解。实际上，它对用户的自主性产生了强力侵蚀，在算法权力所设定的场景之中，用户的自我欲望、认知和决策都不断地被塑造、引导（周丽娜，2023），有些主体意识较强的用户可能会意识到算法对自我认知的侵蚀，而有些用户则在不知不觉间被模塑。同时，由于算法与用户之间的权力不平等，作为场景设置者的算法掌握着全局信息和设计意图，而作为场景参与者的用户则处于信息盲区，这种结构性的不对等使得反抗或协商变得异常困难。

算法权力的多维渗透高效且隐蔽，它通过规则外化将自身意志变为不可挑战的算法规则，再通过时空转化对用户进行“全景敞视”的行为监控，最后通过一个个的场景设置构建了用户“被安排”的现实，使用户在无意识中按照算法预设的路径行动。这三个维度相互交织，共同构成了全方位的算法权力控制样态，对网络生态、用户和社会环境构成了深刻的挑战。

三、算法权力控制下的意识觉醒与互动博弈

经由算法权力通过行为规训、时空控制以及场景设置三种形式的渗透，用户们也不全然是冷漠悬置型反应。在此过程中，他们也有尝试过或明或暗的抗争。但数据产生过程的复杂性乃至算法“黑箱”，往往阻碍了其理解奖惩的内在逻辑，且反抗的对象从有形的管理者（平台管理者）演变为无形的数字技术系统，这无疑也增加了新的难度（贾开和胡凌，2024）。处于被控制地位的用户们只能根据算法系统的漏洞不断想出应对之策，而系统也会为应对出现的新状况而不断升级，循环往复，最终形成一种“用户反抗—系统迭代”的数字缠斗局面。

（一）主动抵抗算法权力规训之获取式策略

获取式抵抗意指，用户主动抵抗来自算法权力的规训，呈现为在抵抗状态上用户自我意识的基本觉醒，其典型表现是进行数据归类与数据干扰。数据归类即用户通过使用多个账号的方式将自我数据分类处理，^①可以按照业务信息、敏感信息、个人信息、组织机构信息、客体信息、系统数据、基础类型、统计信息等进行分类。数据干扰则是用户通过故意点击、故意搜索等方式让自我数据处于无序化的状态（洪杰文和陈嵘伟，2022）。以抖音短视频为例，用户的停留时间、点赞类别、观看次数都会成为算法捕捉的对象，算法将此类数据进行统计分析，最后再反馈给用户，通过一次次优化实现精准推送。但用户若想规避某类信息，可以刻意地划走、点击“不感兴趣”或对不相关的视频信息点赞、过度停留等方式打乱算法结果，以此来改变抖音通过预先设定好的算法推送内容。此种抵抗方式相对容易，无需借助专业的技术工具，仅通过日常操作的刻意调整即可实现，这也使其成为抖音平台中最具普遍性的用户抵抗形式。但同时，这种抵抗行为的效果具有暂时性与有限性，算法可通过对用户后续行为数据的持续采集与重新分析，快速修正被干扰的偏好判断模型，再度实现对用户行为的精准适配，最终陷入“用户干扰—算法修正—再度推送—再度干扰”的

^① 简单来说，数据归类就是把相同内容、相同性质的信息以及要求统一管理的信息集合在一起，而把相异的和需要分别管理的信息区分开来，然后确定各个集合之间的关系，形成一个有条理的分类系统。

博弈循环。

（二）隐形抵抗算法权力监视之防御式回避

防御式回避是用户主动将自己平时在网络上留下的数据隐藏起来，或者刻意阻断自己在网络上呈现出的个人习惯，以此达到规避算法的不当控制。由此可知，防御式战术主要包含数据隐藏与数据阻断两个方面。数据隐藏指用户通过退出登录、不点赞等方式隐藏自己的真实需求；数据阻断指用户通过关闭应用权限的方式阻止算法对相关数据的采集。一般而言，防御式战术贯穿用户使用数字网络的全过程，因此这一战术采用大多是长期性的（洪杰文和陈嵘伟，2022）。以算法对外卖员的行为控制来说，外卖骑手并非完全服从于算法的规训，而是会在熟悉算法的运行规则后利用算法的漏洞，采取各式各样的方法来反抗算法的控制（孙萍，2019）。

（三）集体抵抗算法权力认知塑模之组织化抗议

用户联合起来的集体行动与逆向评价代表了他们抵抗算法控制的组织化努力，是算法权力控制下显著的反抗形式。与用户个体的主动抵抗和被动防御策略不同，集体行动通过劳动者的团结协作解决算法与劳动关系的问题，对平台的算法结构构成了直接挑战。例如，网约车司机通过集体下线的方式来抗议算法不合理的定价；外卖骑手利用社交平台、网络媒体等揭露平台算法对配送时间的不合理压缩（陈龙，2020）。对于网络苛刻的内容审核机制，用户们也会建立互助小组分享算法考核的经验。网络空间中这些分散且多样的行动，实质上都是针对算法不合理控制、为算法的透明化与用户权益的保障做出的集体性反抗。平台算法通过复杂的评价体系对劳动者进行监控和管理，当突破合理边界时，网络用户们就会通过集体协作的方式，将评价武器逆向指向平台，“逆向评价”就是网络用户们发展出的一种独特的反抗机制，即利用平台自身的评价系统进行逆向操作（刘述良，2023）。例如，用户们发现某平台软件存在不合理收费、强制广告插入、泄露个人隐私等侵害权益行为时，通过网络发帖、组建社群、集中举报、集体差评、统一打低分等方式形成协同行动来抵制该平台对用户权益的不合理控制。当用户们逆向评价达到一定规模时，能够有效制衡平台的单方面决策权，从而以适应劳动者的合理诉求来调整平台算法。在组织化抗议过程中，网络用户们逐渐认识到算法权力控制的本质，从无条件信任算法技术中立转向对算法合理性的质疑。与个体化的抵抗不同，网络集体抵抗能够将分散的不满转化为具有影响力的对抗行为，持续时间较长且影响力较大，甚至直接推动平台算法朝着更符合用户利益的方向变革。

四、“控制—反抗”情境下算法权力与用户权利矛盾的显现

算法权力通过规则外化、时空监视、场景设置三个方面完成对用户的基本控制，其中不合理控制激发用户们对自我权益维护的反抗，而不同的反抗形式又变相催生了算法技术的变革与完善。在“控制—反抗”的互动博弈循环之下，也揭露出算法权力与用户权利之间存在现阶段制度上、实践中无法有效进行调节的纷争与矛盾。

（一）社会规则与算法规则核心架构的矛盾

导致算法权力对用户规训的根源是其内部运行架构的不协调与失当。在算法运行的全过程中，以网络平台为代表呈现出“数据化”的单一模式，导致商家与用户等基础性群体忽视自我需求追求高评分、高流量，逐步形成了被算法数据所规训的主体。与传统社会不同，在数字社会中，人们的

主体行动往往是简单而统一的，鼠标与键盘的点击成为常规情形下人的具体行为，但这一具体行为与传统社会中人们做出的包含主观意图的行为存在本质的差别。网络世界中的社会行动往往以一种编码和还原为信息流流动的方式来完成，以信息生产和流动来实现行动目的的主体不仅仅是人类行动者，也有可能是人工智能行动者（陈甬，2021）。传统社会的“行为—责任”规制模式链条也不再适用，取而代之的是“行为—网络—责任”，网络语言不仅是网络世界的工具，甚或某些特定话语本身就是一种社会实践行为（奥斯丁，2012）。当然，网络不是法外之地，也不是全然地在自由生长，国家也正积极通过法律法规、行业准则、第三方力量合作等手段进行治理。但是问题在于，对网络空间的治理不在于政府愿不愿意规制，而在于政府有没有能力规制。网络空间生而自由，政府可以威慑，但网络行为却无法控制；法律可以制定，但其对于网络空间却没有实际意义。网络空间是一个完全不同的社会，那里有约束和管理，但应从下而上建立，而不是通过国家指导来建立（莱斯格，2018）。究其根本是现有大多数规制手段均立足于物理空间的外在规制，并没有进入网络世界算法逻辑的运行内部，是对代码和算法产生的后果进行的规制，而不是对其运行过程或者行动的规制（马长山，2021）。算法行为的构造是技术逻辑的数学架构，是算法系统做出的意思表示。而法律行为有着严格的法律主体要求和内在意思要素的构建，其外在行为有身体的运动，内在包含意思和意识，是精神的一种属性、状态或活动形式（侯东德和张可法，2022）。不一样的规则体系，势必造就不一样的治理方式，于网络空间而言，如何设计内在的基本架构成为对其进行规制无法规避的问题。

（二）用户个人信息保护与算法全景敞视主义的矛盾

算法权力控制下的过度监视侵害了用户的隐私权益。在算法权力压缩的时空情境下，用户对时空的感知超脱了传统的物理限制。网络世界的触角逐渐伸展到物理世界，用户的网络行为会不断延伸与交织，最终转化为影响现实的动态力量。但基于算法权力对时空的控制，借助传感器和移动智能设备，用户所做的每一个看似出于个人意志的选择其实都在算法权力的时空监视之下。用户对算法监视的切身感知，以及对个人隐私防护的反抗意识，是算法治理中的另一矛盾。算法作为确定信息流动的力量，操纵了人类获取的知识（张凌寒，2021）。它在网上监控并挖掘个人留下的信息，进而获得精准推送的能力，将广告信息提供给更需要的用户。从粗放的注意力经济到精准的广告推送，数据在其中扮演了关键作用（陈龙，2020）。算法推荐基于用户历史浏览数据及购买行为对用户进行个性化标签分类，推送其可能感兴趣的直播/商品。如果用户长期对某种风格的商品感兴趣，那么推荐可能不仅仅使用了该用户自身的用户画像，可能还会采用协同过滤的方式，计算与该用户相似的其他用户，并找出被其他用户看过而没有被该用户看过的商品来进行推荐（匡文波和陈小龙，2018）。基于用户推荐算法，系统预测用户未来的行为，从而推荐用户最可能买的产品，而不是最适合的产品。算法通过对用户数据的采集与监视，会对不同个体的行为习惯进行人物画像，为支付意愿高的消费者定高价，而给支付意愿低的消费者打折。这种不动声色的数据采集与人物标签，逐渐使用户陷入一种被监视的恐慌之中，用户对网络的信任也随之陷入低谷。

（三）用户的自主选择权与算法蓄意情境引导的矛盾

算法运行过程中的自主决策变相侵蚀了用户的自主选择权。算法权力利用其机器优势，可以实现对海量数据的高速计算，显然这一过程靠人力无法实现，或者说实现需要的时间及物力资源损耗巨大，因此在这一过程中人类被迫交付计算决策权。更为显著的优势是，与人类不同，利用深度

自主学习系统，算法可以从既往的数据中自主测试和自我改善，甚至可以生产新的知识（Deleuze & Guattari, 2009）。新兴的社会以互联网为中心，社会最基础层面流动的不是物质生产而是信息（张凌寒，2023）。琳琅的桌面和多样的选项看似是充分尊重用户自主选择权下的自由选择，但实际上预设的桌面、推荐的选择、信息的关联度等都是算法权力的引导。例如，在数智化高度普及的社会背景下，餐厅采用扫码点餐本身具有合理性，但是仅提供扫码点餐一种方式，客观上对消费者构成了一定程度的强制，消费者的就餐需求与平台算法系统的强制性使用形成第一层矛盾。消费者在被动接受扫码点餐模式后，又会被要求授权手机号码与位置信息。在该界面虽设置了可供消费者勾选“同意”与“不同意”的选项，看似给予了消费者自主选择的权利，但如若消费者勾选了“不同意”则无法使用该点餐系统，对于一家仅支持系统点餐的餐厅而言，无法使用该点餐系统即意味着无法在该餐厅就餐。当然消费者可以有别的选择，比如换家餐厅就餐，但考虑到时间成本与情绪价值，往往最终选择“同意”授权。在此情境下，“通知—同意”作为继续使用该系统的格式条款之一，并未真实反映消费者的自主意愿，而是在事实上形成了非自愿的授权选择。从过程的反复拉扯及试探，可以直观地看出主导全过程的是出于想就餐/准备就餐的理性权衡所不得不做出的被动妥协。除此之外，大量类似商品/信息的推荐，会使其陷入“商品茧房”之中，很难再接触到其他种类的产品（燕道成和李菲，2020）。这一弊端在依托互联网平台就业的劳动者身上体现得更为明显，平台劳动者看似下载了平台App，实际上他们是下载了一整套塑造主体性的时空框架（黄再胜，2022）。比如，平台所有者会通过感知信息透明度来提高消费者的在线购买意愿，而系统感知风险在一定程度上主导了感知信息透明度对购买意向的影响（Kharitonova et al., 2023）。这些数次妥协的背后是算法脚本被设计者“铭刻”于技术物中，往往没有直接写在技术说明书里，但被强加给技术使用者，引导他们去选择某些行动，同时排斥其他选择（张海柱，2026）。因此，在逐渐让渡决策权的过程中，用户的意识不自主地被算法导入了其所预设的各种情境之中。故而，用户的集体性反抗映射出他们对自我意识自由表达的向往与对算法情境引导下权利侵蚀的抵制，此即构成了算法权力与用户权利的第三重矛盾。无论是选择组织化抗议或是用户认为比较有效的逆向评价，均无法从根本上逃脱来自算法权力刻意的情境化引导。

五、算法权力与用户权利矛盾的调适

“控制—反抗”情境之下凸显出的三组基本矛盾，直接指向了算法权力深层架构、运行程序、权利保障机制的不足。因此，对于上述矛盾的调适，亦需要从这三个维度展开，即如何在优化算法权力内在架构的基础上，强化用户个人信息保护、保障用户自主选择权。

（一）以优化算法设计为基础弱化算法规训

欲打破来自算法权力不合理规训的桎梏，需从底层架构、运算逻辑、模块配置、权限机制四方面加以改造，消解算法固有的规训属性，实现技术层面的反规训。从底层架构来说，网络平台内部的三台架构^①是其主要的核心区块。因此，依据前中台聚焦“眼前”的特征，可以汇集来自平台用户所反馈的信息，针对平台被控制主体，分别创设代表各自群体利益的用户服务反馈意见大数据库，然后根据数据库所反馈的信息，对后台进行实时调整。从算法的运算逻辑来讲，应摒弃唯“流

^① “三台架构”是指组织的前台、中台以及后台，其中，前台直接对应客户，中台是为前台提供大而全、高共用性、高重复使用性的资源配置与赋能服务平台，而后台则是居于幕后对前台中台进行长期的战略指导、基础研发、未来市场培育、企业文化与领导力培养。

量论”“偏好固化”“用户塑模”等设计思路，下调个人历史行为、消费偏好、浏览轨迹等的决策权重，转而将用户自主、公平开放、多元供给作为算法底层运行逻辑。以开放包容的思想，搭建人机协同的制衡框架，摒弃“算法说了算”的单向管控模式，从根源上弱化算法对个体用户的单方面规训。在模块配置方面，应进一步加强信息的风险拦截校验模块，实时识别算法操控、过度诱导、强制捆绑等规训行为，在达到触发条件后（如用户的抗争实践），主动询问用户意向，并进一步修正运算结果，阻断不合理的管控指令。在权限机制方面，应当充分发挥架构中的“算法守门人”角色。“算法守门人”深嵌于平台架构之中，利用其互动的权力机制进行可见性调控，可以通过底层代码实现将抽象的治理意志具象化为可执行的技术实践，以事前的风险识别与事后的数据审核为核心实现对算法权力规制的精准把控。其中，“数据审核”立足于对算法逻辑、算法路径以及算法结果的审核；“风险识别”关注事前的规制，可以将实时行为与历史案例进行毫秒级的特征匹配，并综合行为特征与内容属性对欺诈信息实施自动拦截，针对利用机器人账号、垃圾邮件、网络钓鱼等手段操纵舆论的系统性风险，算法守门人可以通过调用跨平台API实现对恶意主体的联合封堵（张入迁，2026）。

具体到不同数字主体的算法架构设计，对于平台经济体系中的头部生产者而言，一是需要汇聚类似网络直播平台的出错率、关停率、纠纷率、投诉率、举报率等信息，以提请被控制主体合理的注意，达到避免头部生产者在进行生产时误触红线，导致影响个人流量及收入的情况。二是反资本权利所设定的算法引流，在头部生产者生产之前，平台需要披露当前直播隐形流量及其他主播转型趋势，避免盲从与过激。对于“流量大咖”，平台需要披露内在原因、粉丝构成及流量汇聚过程，以规避资本对流量蓄意热捧的现象，影响网络环境。对于平台中的普通劳动者来说，应当贯彻“多劳多得”的收入分配策略，平台分配按在线时长及活跃度来设定，根据距离远近按顺序派单，实时路程调控，按出餐（单）后同路线同时间顺手派单，以实现派单效率、骑手资源利用最大化。对于大众用户来讲，算法推荐导致的信息茧房、认知固化等问题，应当建立以用户需求为导向，尊重用户自主选择权，扫码点单有选择拒绝的权利、使用APP也有个人兴趣爱好的选择权利、数据收集时也有信息不被采集的拒绝权利、不当采集时也有合理有效的反馈通道等。以用户为中心，注重用户体验感与反馈，并据反馈及时予以调整，形成一个循环系统，不断优化算法设计。

（二）以“限定+提示”为核心约束算法监视

对算法过度监视的规制，需破解网络空间算法权力渗透无边界的治理困境。算法监视隐蔽性强、难以约束，用户即便发现自身被监视，往往也缺乏有效的应对手段。究其本质，是算法权力边界模糊、行为提示机制缺失、权利保障缺位三重问题叠加所致。对此，首先，需要限定数据监视采集的范围；其次，监视的过程应当设置提醒；最后，被监视用户应当享有拒绝的权利与救济途径。

1. 限定监视场景与数据采集范围

对算法监视的场景与数据采集的范围进行限定，以避免算法对网络用户权利的侵蚀。首先，对于监视场景来说，能够被强制性监视的内容，应当仅限于涉及国家安全、公共安全、应急管理、重大公共利益等法定情形。以国家安全为例，监视的触发应当以涉及国家安全的敏感词、句子语义匹配、涉密文件/格式、对象等为条件，在达到一定次数后，网络可以进行必要监视。而对于商业用途、用户正常浏览、行为预判等非必要用途，应当限制监视。其次，对于数据的抓取，应当以《个人信息保护法》第5、6条为原则，遵循“最小必要”原则，不得无理由、无限度、无差别地对用

户数据进行采集。以《数据安全法》《网络安全法》等法律法规为刚性约束，禁止对生物识别信息（例如个人人脸、指纹、声音等）、敏感隐私（宗教信仰、金融账户、通讯录等）、画像数据（商业偏好、消费记录、社交关系等）等数据进行采集。

2. 设置数字追踪提示

对于被监视的对象与数据采集的范围，应当设置必要的数字追踪提示。现代技术的一个重要缺陷在于，那些因它们的存在而受到影响的人们对于其设计和运行只有很少甚或没有控制权（温纳，2014）。用户的反抗，一方面是由于算法在决策方面建立了与用户之间的技术鸿沟，另一方面也由于算法的不透明性、对个人数据的过度采集与使用等，对用户的知情权、隐私权等方面带来了挑战。因此，算法在对用户进行监视与数据收集时，应当设置相应的提示义务，根据提示，用户可以选择继续或者禁止算法的追踪行动。建立算法权力与用户之间的信任，实现不同场景下“负责任”的算法（丁晓东，2020）。具体而言，算法的提示系统需要向用户说明并证明：“为了控制该用户实施的A风险（如危害国家安全），算法采取了B控制措施（如强制删除违法内容+封禁账号或者向公安机关报告等），并披露其内在的算法控制逻辑、测试结果和持续监控方案。”反之，对于错误追踪与过度采集的用户与数据应当及时终止，并提供事后说明，说明应包含具体事由、过程及用户救济申诉权力等事项。且对于错误收集的数据，用户享有删除权，可要求算法对该部分数据进行永久性删除。

3. 实现算法权力“可追责性”

约束算法权力过度监控的另一重点是实现以可追责性为核心的约束机制。首先在程序层面，要构建可追溯的责任传导链，赋予AI模型开发者生成内容追溯义务。通过溯源技术在内容创制、发布、传播全流程中植入可验证的标识和证据链，实现对信息来源和处理过程的全程追踪，为内容真实性判断提供技术保障（胡泳和王昱昊，2025）。证据链可以采用区块链存证的方式，记录每次规则调整的技术动因，允许监管部门回溯特定时段内算法立法的数据输入与逻辑推导过程。标注的形式可以采用数字水印技术，^①数字水印技术本质上是一种身份识别和信息隐藏技术（严驰，2025；荣先进等，2025），可以为AI生成的内容添加可追溯标识。在此基础上，还应当持续开发多模态水印植入技术，针对文本、图像、音频和视频等不同媒体类型，研发专门的水印嵌入算法，例如，文本水印^②、图像水印^③、音视频水印^④等。这些技术将确保水印在内容编辑、压缩、转码等过程中的鲁棒性，同时保持对人类感知的“不可见性”。最终形成强制标识与全过程操作记录为核心的透明治理。以此，在制约算法权力过度监控的过程中，更容易明确责任主体，保障用户数据安全。

（三）以分散权治为手段弱化算法自主决策权

以分散权治为手段弱化算法自主决策权，避免算法权力的情景式引导，保障用户自主选择。在网络空间中，知情权是用户行使自主选择权的前提，只有知晓自身所享有的选择权限，用户才能有效行使该项权利。知晓算法权力运行的逻辑首先必须把握算法基于数据的知识挖掘、意义解读与操作控制机制的本质，打开算法的黑箱，厘清数据画像等数据智能应用所隐含的能动性（段伟文，

① 数字水印是指在不影响被保护数字对象价值的前提下，利用其中普遍存在的冗余数据和随机性，把文字、图像、声音、随机序列号等具有可鉴别性的特定信息（如版权信息、作者信息、发行人信息等）嵌入到载体中，从而在真假和虚实间构筑起一道屏障。

② 通过调整词汇选择、句法结构和格式特征进行标识。

③ 利用频域变换和感知哈希算法嵌入标识。

④ 通过时域或频域方式，把标识信息嵌入图像像素或音频信号当中。

2018)。利用区块链等技术,将算法决策的系统逻辑、数据依据及结果生成原理等要素形成不可篡改的分布式账本向有权限的监管机构、行业协会以及使用用户开放验证权限。通过增设算法运行逻辑的解释说明义务,给予用户对其决策进行反思的机会,防止算法决策对用户自主权利的侵蚀(安晋城,2023)。更进一步地,当出现由算法自动化决策错误而导致法律责任时,也能较为迅速地确定责任主体(张恩典,2019)。

明晰算法决策的逻辑后,再对算法权力进行系统性分散,可以有效打破来自算法对用户主体认知的模塑,保障用户自主选择权。如果说算法的分类过程是在“了解”人们的行为,那么算法的决策过程则是在大规模地“塑造”人们的行为(刘河庆和梁玉成,2023)。因此,要想减轻算法权力对用户认知的刻意模塑,首先需要打破少数科技公司或平台企业垄断和控制算法研发与应用的局面。政府监管部门需要积极的政策引导与奖惩机制避免算法权力集中化,强调多元的技术治理方案,保障公民在面对算法决策影响时不单单局限于“同意”的选项,也有“退出”的权利。在算法设计中,更要打破“算法技术中立”的意识导向,能够使用户认识到算法的局限性,从而在技术设计中明确限定算法的适用领域并确保人类主体在需要时能够对算法运行按下“暂停键”。

在系统内部进行算法决策权力分散,利用分布式^①治理理念,将算法权力系统中的关键组件,分为不同的节点。例如,数据输入、模型训练、决策逻辑与结果审计交由技术上相互独立、利益上相互制衡的不同主体负责。其中数据采集、清洗、用户画像等由独立数据部门负责,第三方机构负责模型训练,第三方机构只能就数据部门提供的数据进行训练,无权篡改原始数据,进而操纵数据运行结果,实现数据与算法决策的分离。行业协会制定公平性约束规则,并对平台算法设定强制披露义务。最后由独立的监督机构对算法决策进行实时监测,实现对其程序设定的过程性监督。监督内容包括但不限于:在平台活动中,消费类平台所披露的运行逻辑,是否有嵌入大数据杀熟、信息茧房、饥饿销售等隐含机制误导、诱导用户做出选择的机制;信息类平台是否利用算法技术操控用户偏好、限制信息获取、片面塑造用户认知等,单纯以提升页面浏览量、博取流量为导向开展运营;社交娱乐类平台是否通过算法低俗引流、制造流量焦虑、恶意诱导沉迷、煽动对立情绪,是否滥用社交关系链进行过度营销与隐私暴露;金融服务类平台是否利用算法隐瞒风险、误导风险评级、诱导过度借贷或投资等。一旦触及上述情形,用户及相关群体可以通过反馈,要求监管机关对相关平台进行惩治,责令其改正。通过外部与内部相结合的权力分散机制设置以弱化算法的自主决策权,保障用户权利。于政府而言,能进一步加强对算法权力渗透的管理,增强行政权力对网络空间的控制;于算法权力而言,能够确保其始终在动态的社会共识约束下,服务于公共利益,真正实现“技术向善”的治理目标。

六、结语

进入“三位一体”的算法社会,算法权力的影响日益显著。在服务于用户的同时,也形成了“以技术外化为社会规训、以时空刻画全景敞视的监控、以自主决策重塑用户认知”为代表的渗透局面(周辉,2019)。然而,面对这种控制,数字主体并非完全被动,而是发展出了主动抵抗之获取式战术、被动抵抗之防御式回避以及集体行动的抵抗组织化抗议。这种数字缠斗过程中,可以清

^① 分布式的概念起源于计算机科学领域,主要是指将计算任务、数据存储或者业务功能分散到多台相互独立的计算机(节点)上进而协同完成系统任务,其核心是多台机器协同工作,提高计算效率、可靠性和可扩展性。

晰地看到算法权力治理面临着社会规则与算法规则核心架构、用户的隐私权防护与算法全景敞视的监视、用户的自主决策权与算法情境引导之间的深层矛盾。利用算法内部核心运行机制,形成以源头治理为根基,从内部架构进行优化与调适,弱化算法规训;以“限定+提示”为核心,约束算法权力监控;以分散权治为具体手段,弱化算法自主决策权的全局治理模式。完成从算法权力系统架构的调适到用户隐私的保护,最后到用户决策权、知情权的保障,期望实现算法内部权责明晰、算法运行过程监督制约、网络空间发展规范有序的良好治理生态。

但初步的框架设想需要辅以细化的技术变革以及与之配套的制度保障,这不仅需要政府监管的有形之手,也需要市场竞争的无形之手,更需要社会参与的协同之手。未来算法权力治理的关键在于摆脱简单对抗思维,构建架构合理、技术合规、权力合法的多向度的治理机制,在确保技术创新活力的同时,防范算法权力的过度控制,最终实现技术理性与社会价值的和谐统一,为数字时代用户的隐私权、知情权、决策权等基础权利提供切实保障。

参考文献

- [1] 安晋城.算法透明层次论[J].法学研究,2023(2):52-66.
- [2] 奥斯丁.如何以言行事[M].杨玉成,译.北京:商务印书馆,2012.
- [3] 陈甬.信息行动理论——数字社会时代的社会行动理论探讨[J].社会学评论,2021(5):65-84.
- [4] 陈龙.“数字控制”下的劳动秩序——外卖骑手的劳动控制研究[J].社会学研究,2020(6):113-135.
- [5] 丁晓东.论算法的法律规制[J].中国社会科学,2020(12):138-159.
- [6] 段伟文.数据智能的算法权力及其边界校勘[J].探索与争鸣,2018(10):92-100.
- [7] 福柯.规训与惩罚:监狱的诞生[M].刘北成,杨远婴,译.北京:生活·读书·新知三联书店,2019.
- [8] 高进,黄种兴.数字社会的文化图景:算法推荐与民族志解读[J].学习与实践,2025(5):21-30.
- [9] 洪杰文,陈嵘伟.意识激发与规则想象:用户抵抗算法的战术依归和实践路径[J].新闻与传播研究,2022(8):38-56.
- [10] 侯东德,张可法.算法自动化决策的属性、场域与风险规制[J].学术研究,2022(8):36-44.
- [11] 胡凌.数字架构与法律:互联网的控制与生产机制[M].北京:北京大学出版社,2024.
- [12] 胡泳,王昱昊.技术过程论视角下AI幻觉生成的价值负荷与伦理问题探析[J].南京社会科学,2025(3):84-94.
- [13] 黄再胜.算法控制、“自我剥削”与数字劳动的时空修复——数字资本主义劳动过程的LPT研究[J].教学与研究,2022(11):97-112.
- [14] 贾开,胡凌.合作的互联网[M].北京:商务印书馆,2024.
- [15] 匡文波,陈小龙.新闻推荐算法:问题及优化策略[J].新闻与写作,2018(4):66-70.
- [16] 莱斯格.代码2.0:网络空间中的法律[M].李旭,沈伟伟,译.北京:清华大学出版社,2018.
- [17] 蓝江.从规训社会,到控制社会,再到算法社会——数字时代对德勒兹的《控制社会后记》的超一解读[J].文化艺术研究,2021(4):1-11.
- [18] 刘河庆,梁玉成.透视算法黑箱:数字平台的算法规制与信息推送异质性[J].社会学研究,2023(2):49-71.
- [19] 刘述良.驯化数维坦:反数字支配的复合方案[J].开放时代,2023(5):81-95.
- [20] 马长山.迈向数字社会的法律[M].北京:法律出版社,2021.
- [21] 荣先进,王焱飞,胡东辉.人工智能数字水印测评方法研究与测评平台构建[J].宇航计测技术,2025(2):91-96.
- [22] 孙萍.“算法逻辑”下的数字劳动:一项对平台经济下外卖送餐员的研究[J].思想战线,2019(6):50-57.
- [23] 温纳.自主性技术:作为政治思想主题的失控技术[M].杨海燕,译.北京:北京大学出版社,2014.
- [24] 吴椒军,郭婉儿.人工智能时代算法黑箱的法治化治理[J].科技与法律(中英文),2021(1):19-28.
- [25] 肖红军.算法责任:理论证成、全景画像与治理范式[J].管理世界,2022(4):200-226.
- [26] 严驰.数字水印赋能人工智能生成内容(AIGC)标识化管理的现状与对策[J].科技管理研究,2025(8):181-188.
- [27] 燕道成,李菲.场景·符号·权力:电商直播的视觉景观与价值反思[J].现代传播(中国传媒大学学报),2020(6):124-129.
- [28] 袁曾.算法应当被解释吗?——人工智能“可控制”的治理向度[J].法学论坛,2025(1):130-142.
- [29] 张恩典.大数据时代的算法解释权:背景、逻辑与构造[J].法学论坛,2019(4):152-160.
- [30] 张海柱.技术政治视角下的算法及其风险治理[J].探索,2026(2):90-100.
- [31] 张凌寒.权力之治:人工智能时代的算法规制[M].上海:上海人民出版社,2021.

- [32] 张凌寒.生成式人工智能的法律定位与分层治理[J].现代法学,2023(4):126-141.
- [33] 张入迁.平台数智化进程中的算法守门人:生成逻辑、运行机制与治理向度[J].未来传播,网络首发[2026-05-23]:1-13.
- [34] 郑智航.人工智能算法的伦理危机与法律规制[J].法律科学(西北政法大学学报),2021(1):14-26.
- [35] 周辉.算法权力及其规制[J].法制与社会发展,2019(6):113-126.
- [36] 周丽娜.智媒时代算法推荐对用户自主性的解构与重构——基于规则治理的视角[J].现代传播(中国传媒大学学报),2023(10):69-77.
- [37] Braman, S., *Change of State: Information, Policy, and Power*, Cambridge: The MIT Press, 2006.
- [38] Deleuze, G., and F. Guattari, “Anti-Oedipus: Capitalism and Schizophrenia”, *Telos*, 2009(37): 242-248.
- [39] Kharitonova, Y., N. S. Malik, and T. Yang, “The Legal Issue of Deterrence of Algorithmic Control of Digital Platforms: The Experience of China, the European Union, Russia and India”, *BRICS Law Journal*, 2023, 10(1): 147-170.

【作者简介】须 露：西南政法大学人工智能法学院博士研究生。研究方向：人工智能法、网络法。

The Multidimensional Penetration and Adaptation of Algorithmic Power

XU Lu

(School of Artificial Intelligence and Law, Southwest University of Political Science and Law, Chongqing 401120, China)

Abstract: In the era of the digital economy, algorithms have become a tool for shaping social power structures. Masking their inherent value biases behind a façade of technological rationality, they exert an invisible yet systematic form of control through their externalization as social rules, panopticon-style surveillance and autonomous decision-making. However, digital subjects do not remain entirely compliant; instead, they adopt resistance strategies such as active resistance, passive defense and organized protest, creating a dynamic struggle between technology and users. This process exposes deep-seated contradictions between the core architecture of social regulation and algorithmic rules, the protection of users’ personal information and algorithmic panopticon surveillance, and users’ right to autonomous choice and algorithmic contextual guidance. To address the challenges of governing algorithmic power, it is necessary to understand the core operational mechanisms within algorithms. By optimizing algorithmic design as a foundation, we can implement source-level governance to weaken algorithmic discipline; by limiting surveillance scenarios and the scope of data collection, and by establishing digital tracking notifications, we can impose core constraints on algorithmic surveillance; and by utilizing internal and external collaboration to decentralize power, we can weaken the autonomy of algorithmic decision-making. Ultimately, this will achieve the adaptation of algorithmic power rules, the safeguarding of users’ fundamental rights, and the healthy and orderly development of the platform economy.

Keywords: algorithmic power; core architecture; algorithmic surveillance; digital entanglement; algorithmic regulation

(责任编辑：马莹)