引用格式： 刘娜，施颖，魏鑫喆，等.人工智能时代下消费者权益风险管理研究[J].标准科学，2025(10):27-36.
LIU Na,SHI Ying,WEI Xinzhe,et al. Research on Risk Management of Consumers' Rights and Interests in the Era of Artificial Intelligence [J].Standard Science,2025(10):27-36.

人工智能时代下消费者权益风险管理研究

刘 娜[1] 施 颖[2*] 魏鑫喆[2] 胡心如[2]
〔1. 中国标准化研究院; 2. 中国矿业大学（北京）〕

摘 要：【目的】针对人工智能时代消费者权益面临的系统性风险，探索构建以消费者权益为核心的风险防控框架，助力人工智能技术合规化发展。【方法】运用风险清单法识别算法偏见、隐私泄露、红利分配不均、物理安全受威胁、决策偏差、选择权受限六大核心风险，运用风险矩阵法从可能性和严重性维度进行量化评估，结合国际标准、区域法规的实践经验，分析现有标准体系存在的不足。【结果】根据评估结果显示，隐私与数据泄露、算法偏见、决策偏差为高风险领域，技术红利分配失衡等为中风险领域。现有标准在消费者参与机制、跨域协同实施等方面存在明显不足，约束效能衰减。【结论】需针对风险分级完善标准体系，高风险领域强化强制性技术规范，中风险领域提升标准适配性，通过多元主体协同的闭环机制确保标准落地，实现技术创新与权益保护的动态平衡。

关键词： 人工智能；风险管理；消费者权益；多元治理
DOI编码： 10.3969/j.issn.1674-5698.2025.10.004

Research on Risk Management of Consumers' Rights and Interests in the Era of Artificial Intelligence

LIU Na[1] SHI Ying[2*] WEI Xinzhe[2] HU Xinru[2]
(1.China National Institute of Standardization;2.China University of Mining and Technology /Beijing)

Abstract: [Objective]Aiming at the systemic risks faced by consumers' rights and interests in the era of artificial intelligence, this study explores the construction of a risk prevention and control framework centered on consumers' rights and interests to promote the compliant development of artificial intelligence technology. [Methods] Six core risks, namely algorithmic bias, privacy leakage, uneven distribution of dividends, threatened physical safety, decision-making deviation, and restricted right to choose, are identified using the risk checklist method. The risk matrix method is applied to conduct quantitative assessment from the dimensions of possibility and severity. Combined with the practical experience of international standards

基金项目： 本文是中国标准化研究院基本科研业务费资助项目“代驾服务质量提升关键要素与标准研究项目”（项目编号：602025Y-12513）；北京市教育科学“十四五”规划一般课题“首都高等教育数字化转型能力评价模型和方法研究”（项目编号：3040-0009）研究成果。

作者简介： 刘娜，博士，副研究员，研究方向为服务标准化、消费者保护。
施颖，通信作者，博士，副教授，研究方向为标准系统工程与方法、管理决策理论与方法。
魏鑫喆，本科生，研究方向为标准系统工程与方法、风险管理。
胡心如，博士研究生，研究方向为标准系统工程与方法、管理决策理论与方法。

and regional regulations, it analyzes the deficiencies existing in the current standards system. [Results] The evaluation results show that privacy and data leakage, algorithmic bias, and decision-making deviation are high-risk areas, while the uneven distribution of technological dividends and others are medium-risk areas. The existing standards have obvious deficiencies in consumer participation mechanisms and cross-domain collaborative implementation, resulting in the attenuation of constraint efficiency. [Conclusion] It is necessary to improve the standards system according to risk classification, strengthen mandatory technical specifications in high-risk areas and enhance the adaptability of standards in medium-risk areas. A closed-loop mechanism with multi-subject collaboration should be established to ensure the implementation of standards, so as to achieve a dynamic balance between technological innovation and the protection of rights and interests.

Keywords: artificial intelligence; risk management; consumer rights; multi-actor governance

0　引 言

随着互联网、物联网和云计算等高新技术的迅猛发展，全球数据呈现出指数级、爆炸式增长态势，“一个‘一切都被记录，一切都被分析’的数据化时代的到来，是不可抗拒的”[1]。在这一背景下，人工智能（AI）已逐渐从辅助性工具演变为重塑消费生态的核心驱动力量，显著拓展了服务普惠性的广度和深度。消费者权益通常指消费者在购买、使用商品或接受服务过程中依法享有的权利，包括但不限于知情权、选择权、公平交易权与安全权等。在数字时代，尤其是人工智能广泛应用的背景下，消费者权益的内涵与外延正被深刻重塑：一方面，数据权利、算法透明度等新型权益逐渐凸显；另一方面，传统权益如知情权与自主选择，因算法决策与自动化服务的介入而面临重构。

随着人工智能系统的日益强大，其影响也愈发显著，特别是当人工智能在能力和成本上超越人类时，其应用范围、潜在机遇与风险都将急剧扩大[2]。AI系统以“智能助手”身份嵌入消费决策全流程，既带来高效与个性化体验，也衍生出新型风险，如非法的生物特征数据采集、算法偏见导致的系统性歧视、生成式人工智能所产生的虚假内容误导等。这些新型风险不仅威胁消费者个人信息安全，更对其决策自主性、公平交易及人格尊严造成深层侵蚀。本研究通过识别六大核心风险，运用风险矩阵法进行量化评估，结合国际标准实践与区域法规经验，构建以消费者权益为底层逻辑的风险防控框架，助力人工智能技术合规化发展的路径选择，推动人工智能在规范框架内实现良性发展。

1　基于风险清单的消费者权益风险识别

在消费者权益保护领域，标准化为风险识别提供了系统化的框架、统一的语言和可衡量的尺度，从而将原本主观、模糊、零散的判断转变为客观、清晰、系统的管理过程，不仅为企业自查、监管部门排查提供了权威依据，更推动风险识别从“经验驱动”向“标准驱动”转型，有效覆盖传统识别手段难以触及的新兴风险领域。

从具体实践来看，不同领域的标准化成果已为消费者权益风险识别提供了扎实支撑。在消费品安全风险管理领域，国家标准化管理委员会发布的GB/T 28803.1—2025《消费品安全风险管理 第1部分：导则》具有标志性意义，立足智能化、个性化消费新趋势，明确要求对AI应用等场景下的潜在安全风险实施动态识别；同时扩充危害源类别，将算法歧视、数据泄露等新型风险纳入识别范畴，并细化评估指标体系，为全流程风险识别提供操作指引。在智能消费品这一新兴领域，国家标准化管理委员会批准发布的《智能消费品安全》系列国家标准针对智能家电、可穿戴设备等产品的技术特性，首次将“信息安全风险”“伦理风险”纳入识别体系，明确提出人脸信息非法采集、AI推荐算法误导等风险的识别方法，填补了智能消费品风险识

别的标准空白。

1.1 基于国际标准和问卷调查识别风险

人工智能时代下消费者权益风险在《ISO/IEC 42001 人工智能管理体系白皮书—AI风险治理》中被分为隐私侵犯风险、物理安全风险、虚假内容风险、算法偏差风险、接入不平等风险等类别。

本文采用德尔菲法，通过问卷调查的方式，收集汇总资料，并结合相应文献资料和国际标准，构建基于文献回顾和问卷调查的风险清单（见表1）。

1.2 风险类别

1.2.1 算法偏见与公平性风险

根据ISO/IEC 23894:2023《信息技术—人工智能—风险管理指南》，该风险可定义为人工智能算法在训练或决策过程中，因嵌入历史数据中的结构性不平等或同质化推荐机制，导致输出结果存在系统性偏向，违背公平性原则的风险。

（1）种族偏见风险

美国某医疗AI因训练数据中非裔患者样本不足，导致糖尿病并发症误诊率高达42%，间接造成数百人延误治疗。当训练数据中嵌入结构性不平等时，算法将通过概率模型将这种不平等投射到未来决策中，这种预测性歧视的生成机制，使得算法既成为社会现实的镜像，又成为不平等再生产的工具[5]。人工智能本应成为包容所有人的平台，而非加剧歧视与分裂的工具。随着人工智能技术的不断普及，受到影响的范围越来越广泛，带来的精神与物质层面的损失也不容小觑。

（2）性别偏见风险

“粉红税”由美国经济学家Ian Ayres提出，即在购买质地、用途、款式相似的商品或同类型服务时，女性消费者支付的金钱往往比男性消费者要多的现象。2023年消费者发现欧莱雅天猫旗舰店中，成分、功效、含量基本相同的女士洗面奶比男士洁面乳售价高出近一倍，5名学生因此向法院提起诉讼，最终欧莱雅全额退款。给产品打上女性“专属”等标签，价格会大幅度提升，这不仅损害了女性群体的自身合法权益，还违背了社会公平正义原则，也触碰了男女平等的价值底线。

（3）信息茧房风险

根据消费者偏好定制个性化内容推荐已成

表1 基于文献回顾和问卷调查的风险清单

一级风险	二级风险	指标来源
算法偏见与公平性风险	A1种族偏见风险 A2性别偏见风险 A3信息茧房风险	IEEE 7003—2024《算法偏见考量标准》
隐私与数据泄露风险	A4隐私泄露风险 A5数据滥用风险 A6信息过度采集风险	南方共同市场《关于数字环境中民主和信息完整性的主席声明》 ISO/IEC 29151 ISO/IEC 25059:2023
技术红利分配失衡风险	A7技术接入不平等风险 A8普惠性标准缺失风险	李哲罕[3] 教科文组织《人工智能伦理问题建议书》
物理安全风险	A9人身安全风险 A10非法利用风险	ISO/IEC 27403:2024
决策偏差风险	A11认知风险 A12经济损失风险 A13市场环境破坏风险	ISO/IEC TS 12791:2024东盟《人工智能治理与伦理指南》
选择权受限风险	A14技术垄断风险 A15选择空间压缩风险	李子浩等[4]

为当今算法的主要逻辑，这也是算法偏见的一种表现形式。调查显示，持续推荐同质化信息会使用户在社交媒体上遇到异质化信息的概率仅为5%～8%[6]。此类算法在与消费者拉近联系的同时，也会因同质化内容推荐使消费者被困于“信息孤岛”，放大消费者的“固有偏好”，甚至将偶然偏好固化为长期执念，最终引发冲动消费或过度消费。比如曾因颜值购买过某种特定类型餐具的用户，可能被算法持续推送高颜值但实用性低的同类商品，逐渐形成为颜值买单的偏见，最终囤积大量闲置物品。

1.2.2 隐私与数据泄露风险

根据ISO/IEC 27001《信息安全管理体系》，该风险可定义为人工智能系统在大规模采集、存储、处理用户数据过程中，因数据全周期管控缺陷，导致数据未经授权被访问、使用、披露或传播的风险。随着数字时代的来临，私人领域以一种前所未有的速度进入公共空间，公共领域和私人领域的界限不断坍缩乃至渐趋消融，原本可由隐私主体有效控制的私人信息不可避免地进入公共领域，随时处于被他人“凝视”与“知晓”的状态[7]。

（1）隐私泄露风险

以2023年WPS隐私泄露风波为例，大量消费者反馈使用云文档功能时，其尚未公开的商业计划书、合同草案等私密文件，被无故标记为“敏感文件”，经调查源于WPS引入了人工智能文本识别算法，却因算法逻辑缺陷与权限管理漏洞，导致文档在未经授权的情况下被过度扫描与不当传播。此次事件波及消费者数量高达数百万，不仅引发消费者对办公隐私安全性的强烈担忧，更导致部分企业用户紧急切换办公软件，以规避数据泄露风险。大量隐私信息遭泄露，会增加社会公众的焦虑程度，ISO/IEC 27001的标准内容仍需继续落实，隐私保护力度继续加强。

（2）数据滥用风险

人工智能系统为了进行训练和优化，需要收集和分析大量用户数据，这更增加了数据泄露和滥用的风险。尽管ISO/IEC 42001提及数据安全管理要求，但众多平台仍凭借算法对用户浏览记录、购买偏好、搜索历史等数据进行深度剖析，构建用户画像，从而进行精准营销，影响了消费者正常的购物体验，反映出标准在数据全周期管控中的落实程度不够。

（3）信息过度采集风险

消费者的数字身份被过度采集，其面部识别、生物特征、搜索记录等数据可能在未经本人同意的情况下被追踪和使用，甚至被复制和传播。2023年中央电视台的3·15晚会曝光了keep在获取用户运动数据的同时，暗中收集睡眠记录、地理位置甚至手机相册权限，并将这些非法获取的信息卖给保健品等公司。这不仅严重侵犯了消费者的隐私权，还可能导致消费者遭受经济损失和精神困扰，降低消费者对相关企业的信任度。

1.2.3 技术红利分配失衡风险

根据ISO/IEC 13066《信息技术—辅助技术互操作性规范》中“技术可及性”相关条款，该风险可定义为因不同地区、群体间技术接入条件存在显著差异，导致人工智能技术红利向优势群体倾斜，形成数字鸿沟的风险。

（1）技术接入不平等风险

由于数字化全球进程无法保证人工智能的均衡发展，不同地区、不同群体之间存在技术接入不平等的问题，使得人工智能技术的红利分配失衡，造成相应的信息误差、知识分割、贫富分化，最终导致数字鸿沟的产生。在一些发展中国家或偏远地区，由于网络基础设施落后或缺乏相关技术支持，当地消费者无法享受到技术带来的便利，被排除在人工智能驱动的各项公共服务之外，技术接入的不平等，使得一部分消费者难以参与到数字经济发展中，在现实中巩固与加剧了社会不平等[8]，进一步拉大了不同群体之间的发展差距。

（2）普惠性标准缺失风险

当前缺乏人工智能服务普惠性的强制标准，如农村或不发达地区智能服务覆盖率等指标未被纳入规范，导致技术红利向优势群体倾斜。同时缺乏普惠性标准的人工智能通常以“主流群体需求”为

核心，忽视老年人、残障人士、低收入群体等的特殊需求，加剧了数字鸿沟，违背了社会公平的基本准则，也不利于人工智能技术的全面健康发展。

1.2.4 物理安全风险

根据ISO/IEC 27035《网络安全事件管理指南》，该风险可定义为人工智能与物联网融合应用中，因智能设备的联网决策功能存在漏洞，被非法入侵或滥用，导致设备异常操作并威胁人身安全的风险。

（1）人身安全风险

由于人工智能技术与物联网的深度融合，该技术被整合并广泛应用在人们生活的各个方面，使得传统的物理产品具备了联网和决策能力，智能家居的发展就是典型例子。当设备被恶意入侵时，可能会诱导做出危险行为，引发安全事故。例如，智能家居中的门锁被入侵可能导致陌生人非法进入等安全问题。这些情况不仅会对消费者的生命健康造成直接危害，还会让消费者对人工智能相关产品产生恐惧心理，影响其对新技术的接受和使用，阻碍人工智能技术在日常生活中的进一步推广和应用。

（2）非法利用风险

各种联网设备在运行时所产生的相关数据都存在一定的安全隐患，可能会被企业非法利用，也可能被不法分子恶意入侵。以“MaskPark树洞论坛”事件为例，大量家用摄像头的录像信息被窃取后，不法分子通过筛选、剪辑，将包含用户居家活动、家庭成员面貌甚至私密场景的片段发布至非法平台，以此吸引流量、售卖会员牟利，这些本应作为家庭安防屏障的设备，却因数据安全防线的失守，最终成为窥探隐私的“眼睛”。

1.2.5 决策偏差风险

根据ISO/IEC 42001《信息技术 人工智能管理系统》，该风险可定义为生成式人工智能因具备高逼真性、低成本的文本、图像、音视频生成能力，产生虚假内容并渗透至信息环境，干扰消费者决策判断的风险。

（1）认知风险

随着生成式AI的出现，“AI生产内容”将会成为一种全新的内容生产模式[9]。生成式AI具备生成文本、图像、音频或者视频的能力，该项技术在给人们带来便利的同时，也伴随着巨大的风险。其生成的虚假内容有着极高的逼真性和极低的传播成本，使得虚假内容能够轻易渗透到生活的各个方面。比如通过生成式AI制作产品宣传，存在放大或歪曲产品功效的可能性，使得消费者难以通过常规经验加以辨别，形成对产品的错误认知，根据呈现的宣传效果而进行冲动消费。

（2）经济损失风险

消费者高度依赖外部信息进行决策，而虚假内容的传播会污染信息源，对决策进行干扰，最终使得消费者做出非最优决策，造成经济损失。比如生成式AI可快速伪造专家推荐视频，让冒牌医生对着镜头宣称某款保健品能治愈某种疾病，利用人们迫切治病的心理谋取利益，使消费者不仅承受经济损失，还有可能承受假药带来的副作用风险。这些虚假内容凭借技术手段规避平台审核，进一步模糊了真实与虚假的边界。

（3）市场环境破坏风险

虚假内容的传播会破坏产业生态，降低消费者的信任程度，扰乱公平的市场竞争环境。部分不良商家可能会从功效和呈现效果方面依赖人工智能进行造假从而抢占市场份额，最终导致“劣币驱逐良币”的现象，不利于市场的健康发展和消费者权益的长远保障。

1.2.6 选择权受限风险

根据ISO/IEC 42001《信息技术 人工智能管理系统》中“市场多样性”条款，该风险可定义为人工智能领域因企业构建技术壁垒、挤压市场份额形成寡头垄断，导致市场替代商品、服务减少，消费者选择空间被压缩的风险。

（1）技术垄断风险

人工智能技术呈现的寡头垄断发展趋势，相关企业通过构建技术壁垒、挤压市场份额等方式，降低了新进入者抢占市场份额的可能性，削弱了替代商品或服务的可能性。其核心表现为：头部企业

凭借在人工智能核心技术与数据资源上的先发优势，通过构筑技术壁垒、实施排他性策略等手段排斥竞争，最终形成寡头垄断格局，削弱市场创新活力并压缩消费者选择权。

（2）选择空间压缩风险

以国内短视频平台市场为例，抖音和快手积累了海量用户与数据资源，2024年市场份额总和超过70%，形成了近乎寡头垄断的局面，在此情形下，两大平台为追求商业利益，逐步降低内容审核标准，致使低质量、同质化内容泛滥。同时，平台内广告投放成本也不断攀升，而消费者由于缺乏其他具有同等影响力的平台可供选择，只能被迫接受这些变化。不仅在内容消费上失去了多元化的选择，还在广告侵扰下降低了使用体验，严重侵犯了消费者的自主选择权，抑制市场的创新活力，不利于人工智能技术的持续进步和行业的健康发展。

1.3 风险关联性分析

人工智能时代下的消费者权益风险并非孤立存在，而是通过数据流转、算法决策、市场结构等纽带形成相互传导的动态网络，某一风险的爆发可能成为另一风险的诱因，最终形成“风险叠加效应”，加剧对消费者权益的侵害。

1.3.1 数据滥用链条

消费者隐私泄露导致其用户数据被非法滥用，会直接造成用户画像失真，基于失真画像的算法推荐会进一步强化偏见，最终压缩消费者的有效选择空间。例如，淘宝、京东都曾因滥用用户浏览数据和消费记录来判定消费者消费水平，生成片面画像，导致低收入用户在搜索界面购买商品时持续被推送不知名品牌的劣质商品，形成“消费歧视”（见图1）。

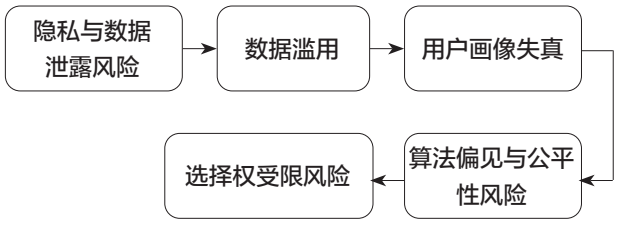

图1 数据滥用链条

1.3.2 垄断强化链条

每个行业的头部企业通过构建技术壁垒加强数据垄断，实施排他性合作进一步排斥竞争，行业进入可能性受限，头部企业的优势进一步固化，技术红利向优势群体倾斜；而中小厂商因缺乏先进的数据与算法资源，难以提供替代的产品与服务，最终使消费者陷入“被动接受”的市场困境（见图2）。

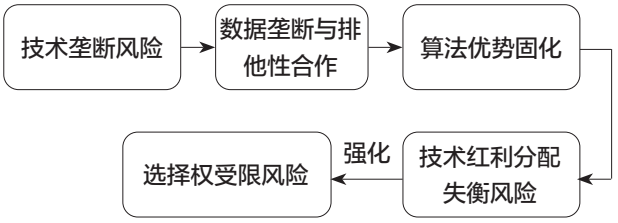

图2 垄断强化链条

1.3.3 虚假信息链条

生成式AI制造的虚假内容流入信息环境后导致信息源污染，消费者因难以辨别真伪而误信虚假信息，并基于错误认知产生相应行为，比如转发包含虚假内容的视频。由于算法偏好性设置，在捕捉到消费者浏览这些被污染的数据后，会强化对同类虚假内容的推荐，导致算法推荐偏差加剧，进而使消费者被长期困在虚假内容的信息茧房中，潜移默化地固化用户的偏见认知，消费者根据虚假内容形成的“生活常识”使用设备时，可能会因错误操作或误开相应权限导致相应物理安全事故的发生，损害自身安全和财产安全（见图3）。

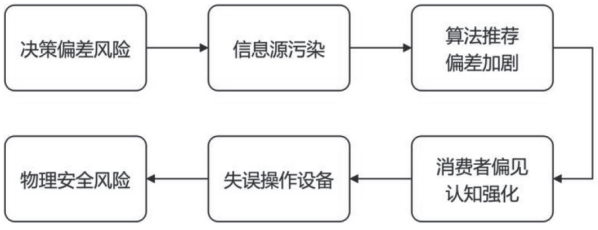

图3 虚假信息链条

2 消费者权益风险评估分析

2.1 风险矩阵评价法

风险矩阵是在项目管理过程中识别项目风险重要性的一种结构性方法，该方法能够对项目风

险的潜在影响进行评估，是一种操作简便且定性分析和定量分析相结合的方法[10]。风险矩阵法首先从可能性和严重性2个角度明确风险维度，可能性是指事件发生的概率，严重性是指事件发生的影响程度和范围，风险发生的可能性和严重性评价标准见表2和表3。

表2 风险发生可能性的说明

风险发生概率范围	风险发生可能性量化值	定义或说明
0~0.5	1	极不可能发生
0.5~1	2	发生的可能性很小
1~20	3	有可能发生
20~50	4	发生的可能性很大
50~100	5	极有可能发生

表3 风险严重性的等级说明

风险影响等级	风险影响量化值	定义或说明
关 键	4~5	一旦风险发生，将导致整个项目失败
严 重	3~4	一旦风险发生，将导致项目的目标指标严重下降
中 度	2~3	一旦风险发生，项目受到中度影响，但项目目标能部分达到
微 小	1~2	一旦风险发生，项目受到轻度影响，但项目目标仍能达到
可忽略	0~1	一旦风险发生，对项目计划没有影响，项目目标能完全达到

在对事件可能性与严重性的评估后，通过在风险矩阵中列举，从而确定风险等级。风险等级的确定见表4。

表4 风险等级对照表

风险可能性量化值 \ 风险影响量化值	1	2	3	4	5
1	低	低	低	低	中
2	低	低	低	中	中
3	低	低	中	中	高
4	低	低	中	高	高
5	低	中	中	高	高

2.2 风险评估结果分析

基于前期梳理的风险清单，邀请5位专家对风险进行评级，专家均拥有多年消费者权益风险管理经验，其专业能力为风险评估结果的可靠性提供了坚实保障。研究采用风险矩阵法对人工智能时代下的消费者权益风险开展系统性分析，相关分析结果如图4所示。

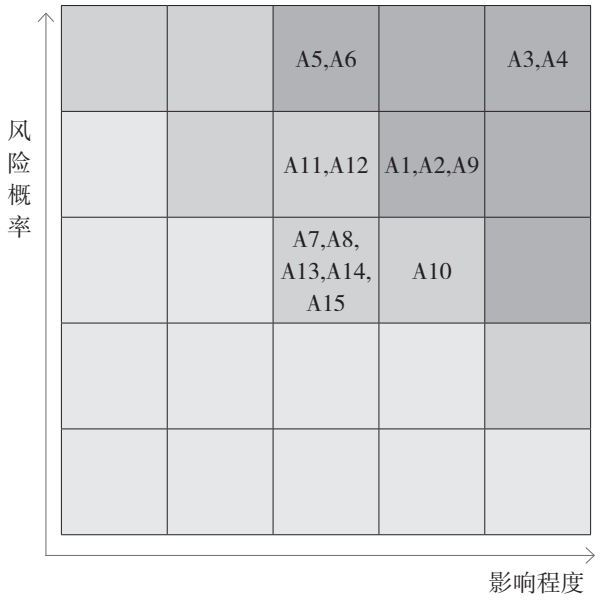

图4 消费者权益风险矩阵图

在风险矩阵得出的结果中，将各类风险按照高、中、低三种优先级进行划分。

（1）高优先级

信息茧房风险（A3）和隐私侵犯风险（A4）属于“高影响—高概率”区间，是消费者权益保护的核心威胁。信息茧房通过算法推荐的内容同质化，加剧认知偏差与社会分化；隐私侵犯则使消费者面临精准诈骗、人格权侵害等风险。种族偏见（A1）、性别偏见（A2）和人身安全风险（A9）位于“中高影响—中高概率”区间，其中种族偏见和性别偏见都属于算法偏见风险，算法偏见对特定群体实施差异化待遇，破坏公平消费生态；人身安全风险聚焦家庭、校园等场景，直接威胁消费者生命健康。数据滥用（A5）与信息过度采集（A6）风险二者同属“隐私与数据泄露风险”，处于“中高影响—中概率”区间，超范围数据采集后，若用于非

法营销等不法途径，将对消费者自主权造成持续性侵害。

（2）中优先级

认知（A11）与经济损失（A12）风险都属于“决策偏差风险”，位于“中影响—中概率”区间。算法诱导非理性消费、虚假信息误导投资决策等行为，使消费者在认知偏差的同时也遭受经济损失。技术接入不平等（A7）与普惠性标准缺失（A8）风险从属“技术红利分配失衡风险”，处于“中影响—低概率”区间，此类风险源于人工智能时代相应的标准不够完善而产生的数字鸿沟现象。非法利用（A10）风险属于“物理安全风险”，位于“中高影响—中概率”区间，表现为智能设备信息被用于非法活动。

（3）低优先级

市场环境破坏属“决策偏差风险”，技术垄断与选择空间压缩为“选择权受限风险”，处于“低影响—低概率”区间，此类风险表现为垄断抑制创新、挤压消费选择。

3　消费者权益风险的应对策略

在参考加拿大《自动决策指令》、ISO“风险等级分类逻辑”等国际经验的基础上，需强化我国《新一代人工智能伦理规范》《中华人民共和国个人信息保护法》的适用，按照风险程度对人工智能侵害消费者权益的行为实施差异化管控，从而精准规避风险，形成更具针对性的风险防控框架。

3.1 风险规避

3.1.1 构建透明化技术治理标准框架

信息茧房风险属于高优先级风险，其核心危害在于算法推荐的同质化内容限制消费者信息获取的广度，加剧认知偏差，应当从算法透明化角度规避该风险，东盟在《人工智能治理与伦理指南》中要求组织披露人工智能系统的使用目的、决策过程及数据来源，使消费者能清晰识别算法是否存在偏见或虚假信息；ISO/IEC25059:2023《软件工程—人工智能系统的质量模型》将“透明度”“可干预性”纳入质量指标，要求开发方披露系统决策逻辑；我国《新一代人工智能伦理规范》明确“算法透明可解释”原则，要求企业对自动化决策的逻辑、依据进行公开。美团此前的算法在配送时间设定等方面存在问题，导致配送超时率、交通违章率、事故发生率上升，为解决这些乱象，落实相关部门算法治理要求，美团在2024年推出八项算法改进举措，包括推进算法公开常态化这种透明化机制，不仅减少了消费者因不知情导致的权益受损，也为风险发生后的责任追溯提供了依据。

3.1.2 构建标准化权利保障框架

隐私侵犯风险对消费者数据主权构成严重威胁，属于高优先级风险。针对生物识别、深度伪造等易引发隐私侵犯、决策操纵的高风险人工智能技术，需通过法规明确禁止其在非必要场景的使用。比如欧盟《人工智能法案》中明确指出对人工智能系统进行约束需要明确界定的方法，并且详细阐释了对高风险技术的管理措施[11]。对生物识别系统在公共场所的使用设置严格场景限制，仅在特定安全保障需求且符合严格程序下才可使用面部识别，充分保障消费者隐私自决权。加拿大《自动决策指令》遵循了按照风险等级进行分类治理的制度逻辑[12]，涉及对健康、安全及基本人权会造成风险的证据，潜在风险的严重程度，已出现风险的性质[13]等多个维度，通过提高准入门槛，避免其未经规范即进入消费市场；我国《个人信息保护法》明确“个人对其个人信息的处理享有知情权、决定权”，将生物识别、行踪轨迹等数据列为“敏感个人信息”，要求“单独同意”。在滴滴数据安全审查案例中，监管部门正是依据《中华人民共和国网络安全法》《中华人民共和国数据安全法》及《中华人民共和国个人信息保护法》，要求其整改数据跨境传输流程，删除非必要收集的用户行踪数据，这一案例印证了国内法规对隐私风险的阻断作用，但标准刚性不足，部分企业对“数据跨境合规”仍存在侥幸心理，需将关键条款转化为强制性国家标准。

同时，信息不对称也是隐私侵犯风险的重要诱

因，可以通过赋予消费者对高风险人工智能交互的拒绝权，从个体层面实现风险规避。联合国教科文组织在《人工智能伦理问题建议书》中从伦理层面为人工智能研发、部署与使用提供指导，要求尊重人权、促进公平正义、保障透明度与可解释性，这些原则构成了保护消费者权益的伦理基石。在消费者权益保护实践中，联合国要求人工智能系统在与用户交互时明确告知身份，同时，确保用户对数据使用和算法决策拥有双重选择权，消费者有权拒绝个性化推荐，避免因过度个性化导致隐私侵犯与信息茧房，提升消费者对人工智能的掌控感。

3.1.3 构建标准化数据管控体系

算法偏见风险的根源往往在于训练数据的结构性失衡，当数据样本无法真实反映社会群体的多元构成，或含有历史积淀的歧视性信息时，算法便会将这种失衡转化为系统性偏见，并通过决策输出持续强化。需以国际标准与区域性法规为框架，构建全链条的偏见预防机制。ISO/IEC 27701与欧盟《通用数据保护条例》所确立的“数据最小化”与“目的限制”原则，为训练数据的范围划定了核心准则，数据需在“满足算法功能需求”与“覆盖多元群体”之间找到平衡。应强制要求训练数据必须包含不同性别、种族、年龄等维度的代表性样本，并通过标准明确各类群体的占比，通过硬性标准避免因样本偏差导致的算法偏见。在数据处理的质量管控层面，ISO/IEC 42001:2023《人工智能管理体系》中“数据质量管控”的要求建立标准化的数据清洗流程，识别并剔除历史数据中隐含的歧视性信息，为算法的公平性奠定坚实基础。

3.1.4 构建智能产品人身安全风险防范标准体系

在当前人工智能与物联网深度融合的背景下，人身安全风险对消费者生命健康构成直接且严峻的威胁。从安全防范角度来看，ISO/IEC 27032—2023《网络安全 互联网安全指南》中指出物理安全规范；我国《人工智能伦理规范》要求“人工智能产品不得危害人身安全”，强调组织需对网络设施和设备实施妥善物理保护，防止未经授权的访问与破坏。在涉及消费者物理接触的关键场景，诸如智能家居领域，强制推行多重安全设计。以智能门锁为例，需同时支持指纹、密码、应急钥匙三种及以上开锁方式，为用户提供多元安全的解锁途径。并且，当其中任一开锁方式出现故障时，智能门锁系统应依据预设程序，自动触发与公安系统联动的报警机制，迅速联系物业管理部门进行紧急处理，最大程度保障用户被困或遭遇危险时能及时获救。对于未达到上述安全设计标准的产品，应坚决禁止此类产品流入消费市场，从源头上消除可能危害消费者人身安全的隐患，为消费者营造安全可靠的产品使用环境。

3.2 风险降低

3.2.1 构建标准化技术接入平等与数字普惠风险防控体系

技术接入不平等与普惠性标准缺失所引发的风险，长期来看会深刻制约消费公平的实现。这种风险的核心在于，人工智能技术红利的分配不均，导致特殊群体被排除在外，形成数字鸿沟。联合国教科文组织《人工智能伦理问题建议书》中“技术应服务于全人类”的原则，明确企业责任边界；ISO/IEC 30141《信息技术—数字包容指南》强调“数字环境的可访问性”与“技术普惠原则”，要求通过标准化设计确保不同群体能平等获取数字服务；我国《数字中国建设整体布局规划》提出要“构建普惠便捷的数字社会”，打造智慧便民生活圈、新型数字消费业态、面向未来的智能化沉浸式服务体验，围绕这一目标，可以通过“数字基建共享计划”将偏远地区5G基站、智能终端纳入公共服务采购，确保每笔投入都能最大化覆盖未接入群体，逐步缩小数字消费差距，让人工智能技术真正成为惠及全人类的工具。

3.2.2 构建标准化全周期风险动态监测体系

通过“适度容忍、精准干预”实现风险的早发现与早控制，避免小风险累积演变为系统性威胁，此类风险当前虽影响范围有限，但若放任其发展可能破坏市场生态平衡，因此需依托制度化监管构建全周期防控网络，形成覆盖企业数据使用、算法行为、市场竞争策略的动态监测体系。我国《中

华人民共和国反垄断法》第二十二条禁止“滥用市场支配地位”，重点核查数据垄断等风险点，通过行业官网向社会公开，既强化企业自律，又保障消费者的知情权与选择权。当监测数据显示企业市场占有率超过60%，且存在“挤压竞品”的实质性行为时，针对垄断核心环节实施靶向干预，在美团“二选一”案例中，监管部门依据“市场占有率超60%+实质性挤压竞品”标准，要求其解除排他性合作。既避免“一刀切”对市场效率的损害，又能精准打破垄断壁垒，恢复市场竞争性，为消费者权益提供持续且稳定的制度保障。

4 结 论

本研究通过风险识别与评估揭示，人工智能时代消费者权益面临多重风险交织挑战：隐私与数据泄露、算法偏见等风险因高可能性与高严重性成为核心威胁，技术红利分配失衡、物理安全等中风险领域则因标准细化不足持续存在隐患。现有标准体系存在结构性短板——国际通用标准与区域规范在消费者权益保护的针对性上衔接不足，跨主体协同实施机制不完善，导致标准对风险的约束效能衰减。

完善消费者权益保护的标准体系，需紧扣风险分级特征精准发力。对高风险领域，应强化强制性标准建设，将算法透明度、数据安全等要求转化为可操作的技术规范；对中风险领域，需增强标准与实际权益诉求的适配性。同时，需建立标准实施的闭环机制，整合企业自检、政府监管与消费者反馈，确保标准从制定到落地全链条有效运行，最终实现人工智能技术创新与消费者权益保护的动态平衡。

参考文献

[1] 周涛.为数据而生:大数据创新实践[J].中国商界,2016(6):123.

[2] 黄丹.人工智能快速发展下的极端风险管理[J].科技中国,2025(5):92–96.

[3] 李哲罕.社会再生产视角下的数字鸿沟问题：基于布尔迪厄理论的考察[J].求索,2025(4):34–40.

[4] 李子浩,王迎春.科技脱钩下中国生成式AI安全风险识别与应对[J/OL].科学学研究,1–16[2025–08–01].https://doi.org/10.16192/j.cnki.1003–2053.20250722.001.

[5] 张涛.自动化系统中算法偏见的法律规制[J].大连理工大学学报(社会科学版),2020,41(4):92–102.

[6] 张开平. 数字时代的政治传播：理论重构、议题革新与范式转向[J]. 政治学研究，2023(5): 193–206，212.

[7] 王帆宇,孙浩然.大数据时代隐私保护的现实问题与实践路径[J/OL].西华师范大学学报(哲学社会科学版),1–12[2025–07–05].https://doi.org/10.16246/j.cnki.51–1674/c.20250403.002.

[8] Nell Selwyn,Reconsidering Political and Popular Understandings of the Digital Divide[J]. New Media & Socie–ty, 2004,6(3):341–362.

[9] 陈永伟.超越ChatGPT：生成式AI的机遇、风险与挑战[J].山东大学学报(哲学社会科学版),2023(3):127–143.

[10] 常虹,高云莉.风险矩阵方法在工程项目风险管理中的应用[J].工业技术经济,2007(11):134–137.

[11] EU. Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain unionlegislative acts[R]. 2024.

[12] 陈少威,杨涛,贾开.比较政策研究视野下全球人工智能治理模式的差异、共识与改革启示[J].中国行政管理,2024,40(12):15–24.

[13] Canada Government. The Artificial Intelli–gence and Data Act（AIDA）–Companion docu–ment[EB/OL].[2025–04–22].https://ised–isde.canada.ca/site/innovation–better–canada/en/artificial–intelli–gence–and–data–act–aida–companion–document.