引用格式： 刘娜，施颖，胡心如，等.消费者权益保障视角下的人工智能治理标准化路径研究[J].标准科学，2025(11):16-26.
LIU Na,SHI Ying,HU Xinru,et al. Research on Standardization Pathways for AI Governance from the Perspective of Consumer Rights Protection [J].Standard Science,2025(11):16-26.

消费者权益保障视角下的人工智能治理标准化路径研究

刘 娜[1]　施 颖[2*]　胡心如[2]　郝素利[2]

〔1. 中国标准化研究院；2. 中国矿业大学（北京）〕

摘　要：【目的】针对人工智能全球化应用中消费者主体性缺失、隐私安全风险及跨国监管碎片化等问题，构建以消费者权益保障为核心的全球AI治理新范式，实现技术发展与公平治理的平衡。【方法】采用文献分析与案例比较法，系统梳理ISO等国际标准体系及中国、美国、加拿大三国的政策实践；基于双重效应理论，深入剖析人工智能给消费者权益带来的机遇与潜在风险；在此基础上整合现有国际治理框架，探索标准化协同路径的构建逻辑。【结果】研究发现，人工智能在赋能消费者福祉提升的同时，数据滥用等问题加剧了消费者隐私安全与权益公平性危机；各国人工智能治理模式因技术发展阶段、法律体系差异而呈现不同特点，跨国监管协作的缺失导致治理效能不足，亟须强化国际协作；据此提出四维标准化路径，具体包括整合国际规范共识转化为技术标准、建立全流程风险治理框架、强化数据主权保障机制、构建“认证—监管”协同联动机制。【结论】全球人工智能治理需依托多元主体共治模式与区域规则有效衔接，将消费者权益保障嵌入标准设计全流程，通过技术规范与法律责任的闭环衔接，最终实现技术创新与社会公正的协同发展。

关键词： 人工智能；消费者权益；标准化路径；全球治理

DOI编码： 10.3969/j.issn.1674-5698.2025.11.002

Research on Standardization Pathways for AI Governance from the Perspective of Consumer Rights Protection

LIU Na[1]　SHI Ying[2*]　HU Xinru[2]　Hao Suli[2]

(1. China National Institute of Standardization; 2. China University of Mining and Technology, Beijing)

Abstract: [Objective] This study aims to address key challenges in the global application of artificial intelligence (AI)—such as the erosion of consumer autonomy, privacy and security risks, and fragmented transnational regulation—by proposing a new global AI governance paradigm centered on the protection of consumer rights, so as to balance technological advancement with equitable governance. [Methods] Through literature analysis and comparative case studies, this research

基金项目： 本文受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）；航空科学基金项目“航空人工智能标准化体系研究”（项目编号：2022Z064004001）资助。

作者简介： 刘娜，博士，副研究员，研究方向为服务标准化、消费者保护研究。
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。
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。
郝素利，博士，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。

systematically examines international standards systems (e.g., ISO) and policy practices in China, the United States, and Canada. Grounded in the dual-effect theory, it further analyzes the opportunities and potential risks that AI poses to consumer rights. On this basis, an integrated international governance framework is constructed to explore the logic of building a coordinated standardization pathway. [Results] The study reveals that while AI enhances consumer well-being, issues such as data misuse exacerbate crises related to consumer privacy, security, and equity. Moreover, AI governance models vary across countries due to differences in technological development stages and legal systems, and the lack of cross-border regulatory coordination undermines governance effectiveness, highlighting the urgent need for international collaboration. Accordingly, a four-dimensional standardization pathway is proposed, including: integrating international normative consensus into technical standards; establishing a full-process risk governance framework; strengthening data sovereignty safeguards; and building a coordinated "certification–supervision" mechanism. [Conclusion] Effective global AI governance should be based on a multi-stakeholder co-governance model that aligns regional rules and embeds consumer rights protection across the entire standard-design process. By forming a closed loop between technical norms and legal accountability, the coordinated development of technological innovation and social justice can ultimately be achieved.

Keywords: artificial intelligence; consumer rights; standardization path; global governance

0 引言

在数字化浪潮深度演进的背景下，人工智能技术的快速发展和广泛应用正在重塑全球消费生态，消费者对AI技术的依赖程度与日俱增[1]。然而，技术迭代过程中消费者主体性的缺失、隐私保护与算法公平等问题的凸显，以及跨国监管协调的困境，使得人工智能治理面临严峻挑战。这些矛盾不仅体现在技术发展速度与治理滞后的脱节上，更反映在全球化背景下监管碎片化带来的系统性风险中，亟须构建以消费者权益保障为核心的治理新范式[2]。

近年来，在国内和国际层面，公私合作型的人工智能标准举措激增，反映了标准在人工智能开发、应用及其规制中的重要性[3]。本研究聚焦人工智能治理的双重效应、协同机制与标准化路径，旨在通过梳理国际组织的治理实践，探索标准与立法的互动范式，最终形成兼顾技术发展与社会公正的治理方案。本研究采用文献分析与案例比较方法，系统考察国际政策与各国监管经验，致力于摆脱消费者赋权困境、弥合跨国监管裂隙，并通过风险全周期管理与包容性设计规范，推动人工智能从效率工具向权益保障载体的转型，为构建全球人工智能治理体系提供理论支撑和实践指引。

1 消费者权益保障视角下的人工智能治理概述

1.1 概念界定

消费者权益是指消费者为生活消费需要购买、使用商品或接受服务时依法享有的权利及相关利益，以《中华人民共和国消费者权益保护法》等法律法规为依据，主要包括安全保障权、知悉真情权、自主选择权、公平交易权、依法求偿权、求教获知权、依法结社权、维护尊严权和监督批评权等法定权利，既涵盖权利本身，也包含权利衍生的合法利益，旨在通过法律保障平衡消费者市场弱势地位，维护交易公平与市场秩序，促进市场经济健康有序发展。依据ISO的界定，人工智能涉及一系列技术与方法，旨在通过计算机技术模拟人类智能，其技术范畴包括深度学习、自然语言处理、计算机视觉等众多领域。AI技术赋予机器感知、理解、学习、推理、分析及决策等能力，使其能够执行传统上需依赖人类智能的任务。ISO定义中的数据驱动的学习与决策能力、人类指导作用等核心要素，直接关联消费者数据主权

（数据收集需知情同意）与算法可解释性（保障知情权），是AI治理标准化需锚定的技术特性。

1.2 消费者权益保障视角下的人工智能治理研究现状

随着人工智能技术在消费领域的深度渗透，隐私泄露、数据滥用、算法失信等风险对传统消费者权益保障体系构成新挑战[4]，推动人工智能治理研究成为维护市场公平秩序与社会公正价值的关键议题。其中数据隐私争端与算法歧视成为社会关注的核心焦点[5]，而消费者行为数据与企业人工智能算法通过“采集—解析—供给升级—行为强化”形成的闭环正反馈循环，进一步凸显了消费者权益保障在技术应用场景中的复杂性[6]。现有研究围绕具体领域的权益保护路径展开探索：在金融领域，学者针对互联网平台在格式条款设置、信息收集使用及营销宣传等环节存在的问题，提出构建系统性的金融消费者保护法律体系[7]；在人工智能模型训练环节，鉴于海量隐私数据训练导致的泄露与滥用风险，相关研究主张将隐私风险应对策略从个体防御模式转向整体性保护框架[8-9]；在医疗数据分析领域，研究强调通过推广数据匿名化技术、完善法律法规与行业标准及构建协同治理机制，强化患者隐私权保障[10]；在平台用户知识隐私保护方面，有研究提出构建知识隐私友好型协同治理框架，通过层级化规则体系、技术架构嵌入保护理念及综合性监管机制，制衡平台支配地位并实施系统性规制[11]。当前研究已初步识别人工智能应用中消费者权益保障的具体风险场景，并在细分领域探索了有针对性的保护路径，但尚未形成覆盖全链条、跨领域的系统性标准化治理框架。

1.3 ISO主导的人工智能标准体系建设

为了推动人工智能技术的健康发展，ISO成立了专门的技术委员会（ISO/IEC JTC 1/SC 42），负责制定人工智能领域的国际标准，规范缺乏监管、不受约束的人工智能技术，建立统一的人工智能标准框架。目前，ISO制定的人工智能国际标准已形成覆盖全生命周期的技术治理体系，其核心内容聚焦于三大层面：（1）统一基础规范，通过术语标准（如ISO/IEC 22989）明确定义人工智能核心概念，为全球协作提供共通语言；（2）构建技术框架，建立机器学习系统通用开发框架与质量管理模型，从系统设计、性能评估到风险控制形成闭环；（3）整合管理体系，以管理体系标准为核心，贯穿风险评估、责任分配及运行监控，确保人工智能开发部署的安全性与合规性，并持续拓展新兴领域，共同形成“术语—技术—管理—创新”四维协同的标准生态。当前，一些著名的标准组织已发布及正在制定的人工智能标准约400项[12]，表1为国际标准化组织制定的部分人工智能国际标准，覆盖“术语—技术—管理—创新”四维体系的关键维度，直接关联消费者权益保障的核心环节，如风险治理、质量控制、人机交互透明度等。

2 消费者权益视角下的人工智能双重效应分析

2.1 机遇维度

2.1.1 普惠服务升级

（1）推动精准医疗范式革新

在信息技术快速发展的时代，人工智能在医疗领域的广泛应用为医疗数据分析提供了前所未有的机遇[13]。基于机器学习和深度学习等技术的智能算法，能够从海量数据中提取潜在规律，推动个性化医疗发展[14]。在临床诊断中，AI通过辅助病理识别与个性化治疗方案生成，赋能患者充分知悉病情信息并参与治疗决策，拓展了患者在医疗服务中的知情权与自主选择权；在药物研发层面，借助智能解析加速靶向药物开发，不仅降低医疗边际成本，更通过提升优质医疗资源的可及性，使更广泛的消费群体公平享有医疗服务，体现了对消费者公平交易权的保障。

（2）构建智慧农业保障体系

人工智能驱动的智能传感网络、无人机巡检及农业大数据分析系统，实现了农业生产的数据驱动精准管理，显著提升了土地产出效率。在作物

表1 国际标准组织制定的人工智能国际标准

序号	国际标准名称	标准内容	与消费者权益保障的关联
1	ISO/IEC 42001:2023《信息技术–人工智能–管理体系》	从管理体系角度，为人工智能领域合规与风险管理提供全面管理框架，涵盖风险评估、治理架构、责任分配、资源支持、运行管理、绩效评价等规范。该框架可助企业有效管理人工智能系统开发、部署和运营，确保其安全可靠应用	通过建立全流程风险管理框架，明确企业在AI应用中的责任边界，直接保障消费者在数据安全、服务可靠性等方面的权益
2	ISO/IEC 25059:2023《软件工程–系统和软件质量要求与评估–人工智能系统的质量模型》	建立人工智能系统质量模型，引入功能适应性、用户可控性等新概念，为评估系统质量提供多维度指标，帮助开发者和消费者全面了解系统性能与潜在风险	通过规范AI系统的质量评估维度，确保消费者能够获得可控、可靠的服务，强化对自主选择权与公平交易权的技术支撑
3	ISO/IEC 22989:2022《信息技术–人工智能–概念和术语》	该标准涵盖人工智能概念和术语，制定术语并描述与人工智能系统相关概念，涉及广泛技术，包含100多个人工智能常用术语，如透明度、可解释性等，可作为专家开发可互操作性及应用程序、系统、标准和人工智能应用使用指南的基础	通过统一“透明度”“可解释性”等关键术语，为消费者理解AI系统运行逻辑、行使知情权提供基础语义保障
4	ISO/IEC 23053:2022《使用机器学习的人工智能系统框架》	该标准中的机器学习框架解释系统组件及其在人工智能生态系统中的功能，适用于各类组织，包括正在实施或使用人工智能系统的公私企业、政府实体和非营利组织，为基于机器学习（ML）的人工智能系统提供通用框架，指导其设计、开发、部署和管理	通过规范机器学习系统的全生命周期管理，减少算法偏见与数据滥用风险，间接保障消费者的公平交易权与隐私权
5	ISO/IEC PWI 42109《信息技术–人工智能–人机协作》	聚焦人机协作这一新兴领域，研究如何优化人机交互过程，提高协作效率和效果	通过优化人机交互机制，确保消费者在AI服务中保留必要的人工干预权，维护其自主选择权与权益救济的可行性

育种领域，通过智能筛析培育优质品种，为消费者提供安全稳定且价格普惠的食品供给。这种方式既保障了消费者对食品质量安全的知情权，又通过技术优化了实现食品价格可控性，使不同收入群体均能获得质价相符的商品，维护了消费者的公平交易权。

（3）重塑消费服务交互模式

依托深度学习模型的个性化推荐系统，使商品与服务供给精准匹配消费者偏好，既降低信息搜寻成本，又通过充分披露商品信息保障消费者知情权，同时赋予消费者更丰富的选择空间以实现自主选择权。生成式人工智能客服系统突破时空限制，通过7×24小时全时域交互支持，实现消费者诉求的即时响应与复杂问题的动态解析，通过提升服务可及性，为消费者行使求偿权奠定基础，同时优化消费体验以保障其公平交易后的权益延伸。

2.1.2 信息效率与可持续发展赋能

（1）驱动生产力范式重构

人工智能自动化工具通过承接重复性劳动任务，推动人力资源向高附加值领域转移，不仅提升全要素生产率，更通过企业级AI应用降低运营成本，最终转化为更优质的消费端产品与服务供给。从消费者权益视角出发，企业将技术红利转化为质优价廉的商品与服务，实质是通过提升产品性价比保障消费者的公平交易权，同时使消费者能以更低成本获取更优质的商品，拓展了其自主选择的空间与质量边界。

（2）构建智能决策支持系统

基于搜索引擎优化与自然语言处理技术的人工智能系统，有效整合碎片化信息生态，为消费者

提供实时化、精准化的决策支持框架，通过降低信息搜索成本与认知负荷，使消费者能够便捷获取全面、真实的商品与服务信息，既强化了消费者的知情权，又为其科学决策提供依据，从而优化消费决策的科学性与时效性，通过提升信息效率保障消费者自主选择权。

（3）赋能可持续发展治理机制

人工智能的数据聚合与预测分析能力在应对气候变化等全球性挑战中发挥核心作用，通过极端气候事件精准预警强化防灾减灾效能，直接关联消费者的生命财产安全保障权；依托环境承诺履行度动态监测遏制绿色漂洗行为，确保消费者不被虚假环保宣传误导，维护其对商品“环境属性”的知悉真情权与公平交易权。通过技术赋能可持续发展，最终将环境效益转化为消费者的长远权益保障，与联合国可持续发展目标中“确保所有人的可持续未来”的核心诉求相契合。

2.2 风险维度

2.2.1 数据治理与隐私安全风险

在人工智能新时代背景下，消费者权益保护面临诸多新挑战[15]。人工智能系统对海量消费者数据的依赖性，直接威胁消费者的隐私权与安全保障权：（1）生物特征、行为轨迹等敏感信息被过度采集且缺乏有效知情同意机制，实质是对消费者隐私权的侵害；（2）数据泄露事件频发助长黑产交易，形成消费者数字身份安全威胁，个人敏感数据的泄露、非法提供或滥用可能危害人身和财产安全，直接违反安全保障权[16]；（3）物联网设备智能化带来的物理安全漏洞，使家庭安防系统、智能家电等成为网络攻击入口，进一步加剧消费者在物理空间与数字空间的双重安全风险，凸显技术应用中个人数据主权与企业商业利益的结构性矛盾对基本权益的冲击。

2.2.2 算法公平性与信息可信度风险

在算法公平性层面，训练数据的结构性偏差导致系统输出存在种族、性别及残障歧视，形成对特定群体的系统性排斥，算法偏见使部分消费者被剥夺平等的消费机会，既违反公平交易权，又侵害其人格尊严不受侵犯的权利；在信息可信度层面，生成式AI的幻觉效应产生虚假内容，如法律案例伪造、医疗建议错误等，直接误导消费者决策，本质是对消费者知情权的破坏。这两方面共同源于机器学习模型的可解释性缺陷与验证机制缺失，加剧了权益受损的隐蔽性与救济难度。

2.2.3 技术垄断与社会外部性风险

人工智能发展的负外部效应间接侵蚀消费者核心权益，在产业层面形成“数据—算力—资本”三位一体的垄断格局，压缩中小企业生存空间，导致消费者可选择的商品与服务种类减少、议价能力弱化，最终限制了消费者的自主选择权与公平交易权；社会层面加剧造成的“数字鸿沟”使全球26亿未联网人口被排除在AI公共服务体系之外，造成不同群体在消费资源获取上的不平等，违背公平交易权的核心内涵；环境与就业维度的负效应虽不直接指向消费行为，但其引发的生态压力与劳动力市场波动，最终会通过商品价格波动、服务质量下降等传导至消费端，间接损害消费者的长远权益。

3 消费者权益视角下的人工智能监管机制分析

3.1 国家层面治理模式比较

3.1.1 中国政策与监管动态

中国对人工智能的治理秉持“发展与安全并重”的理念，积极构建“共商共建共享”治理格局。发展初期，中国的人工智能治理以促进人工智能产业发展为重点，通过政策引导和资金支持，推动人工智能技术在各领域的应用与创新。中国发布的《国务院关于积极推进“互联网+”行动的指导意见》等文件，旨在推动人工智能与各产业的融合，促进人工智能产业的快速发展，提升其在经济发展中的地位。

随着技术的快速发展，人工智能的风险逐渐显现，中国开始强调安全保障，注重防范人工智能可能带来的风险，确保人工智能技术的安全、可

靠、可控发展。《生成式人工智能服务管理暂行办法》要求算法备案、内容审核，明确人工智能生成内容的知识产权归属；《中华人民共和国数据安全法》的实施明确了数据处理者的安全保护义务，规范数据出境活动，确保数据安全；《中华人民共和国个人信息保护法》聚焦个人信息保护，对个人信息的收集、使用、存储等环节进行严格规范，保障公民的个人信息权益。

中国的人工智能治理注重多元主体参与，政府发挥引导作用，同时充分调动企业、科研机构、社会组织等各方力量，形成协同治理合力。政府负责制定统一的战略规划和政策法规，为人工智能发展提供明确的方向和规范；企业在政府引导下，积极投入技术研发和创新，加强自律并遵守相关法规和伦理准则。

3.1.2 美国政策与监管动态

美国人工智能治理模式采取“轻监管”策略，强调市场主导与企业自我规制。在美国人工智能发展过程中，政府的人工智能治理以市场驱动和技术创新为核心，强调通过自由市场经济和技术创新来推动人工智能的发展[17]。这有利于激发企业的创新活力，推动人工智能技术的快速发展，但也会引发深度伪造滥用、数据垄断等问题。

2023年，美国拜登政府颁布了《关于安全、可靠、可信地开发和使用人工智能的行政命令》，旨在加大人工智能领域的监管力度，提出“安全有效原则”，并设立高级别的监管机构，明确重点监管核心企业，然而因这些措施缺乏强制性，导致治理实践中行业自治仍占据主导地位，政府监管力度相对较弱。特朗普再次执政后，废除相关行政令，承诺加大对人工智能技术的投资并减少监管。基于此，预计在可预见的未来，美国将倾向于采取一种“放任”型的人工智能治理模式。

美国的政治组织形式为联邦制，这使得决策过程多中心化，联邦政府与各州拥有独立立法权，各联邦机构也具备一定自主权，为人工智能治理体系带来了高度的灵活性与适应性，能够迅速响应环境变化，满足不同地域和行业的特殊需求[18]。但分散化的立法和监管模式也导致人工智能治理和标准制定过程中存在一定的混乱与重复，增加企业的合规成本，并造成行政资源浪费。

3.1.3 加拿大政策与监管动态

加拿大政府在人工智能政策制定方面，始终注重平衡人工智能创新与安全，在创新激励上，持续加大资金投入，2024年联邦预算拨款24亿加元发展人工智能。在安全保障方面，加拿大发布针对政府使用人工智能的《自动决策指令》，要求运用标准化算法影响评估工具来确定系统风险，确保政府使用人工智能时履行风险适配义务。2023年，加拿大发布的《可访问且公平的人工智能—技术指南》，迈出了制定人工智能技术标准的重要一步，该指南关注人工智能技术的公平性和可及性，确保不同群体都能受益于人工智能技术。加拿大召集关键行业参与者制定《关于负责任地开发和管理先进生成式人工智能系统的自愿行为准则》，为加拿大的相关企业提供了一套通用的临时标准，规范了生成式人工智能安全、负责任的使用。

加拿大紧跟国际趋势，对人工智能的风险进行分级分类治理[19]。对于医疗、金融、交通等关键领域的高风险应用，实施严格的评估、审查和监督机制，确保其安全性与可靠性；对于低风险应用则营造相对宽松的监管环境，以激发创新活力。

3.1.4 中、美、加三国治理模式对比分析

中、美、加三国在消费者权益视角下的人工智能监管机制呈现显著差异，中国秉持“发展与安全并重”理念，以政府引导构建多元主体协同治理格局，通过《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等法律法规，从数据收集、使用到出境全流程规范人工智能应用，强化对消费者个人信息权益的刚性保障；美国采取“轻监管”策略，以市场主导和企业自我规制为核心，虽曾出台行政令试图加强监管，但因缺乏强制性，加上联邦与州的分散化立法导致治理混乱，这使得消费者权益保护更多依赖行业自治，存在数据垄断、深度伪造等风险隐患；加拿大则注重创新与安全的平衡，通过资金投入激励技术发展，同时发

表2 中、美、加三国治理模式对比分析

维度	中国	美国	加拿大
治理理念	发展与安全并重	市场主导，轻监管	创新与安全平衡
消费者权益侧重	数据隐私、多元协同	企业自律，救济渠道依赖诉讼	分级治理，高风险领域强保护
标准化路径	政府引导+行业标准	自愿性标准，市场驱动	政策与标准协同

布《自动决策指令》《可访问且公平的人工智能—技术指南》等文件，对高风险领域实施严格评估监督、对低风险领域保持宽松环境，以分级分类治理保障不同群体消费者的公平受益权。三国治理模式因理念、制度设计差异，在消费者权益保障的力度、方式及效果上形成鲜明对比，反映出技术发展与权益保护平衡路径的多样性，见表2。

3.2 国际组织协同机制

3.2.1 联合国系统内跨机构治理框架协同

联合国秘书长于2022年9月任命了一位联合国技术特使，旨在推动制定联合国全球数字契约，从而推动构建全球数字技术与人工智能治理的综合性框架，形成包容、开放、安全、可持续的人工智能系统，见图1。2023年10月成立的联合国人工智能咨询机构发布了《为人类治理人工智能》最终报告，提出了人工智能治理系统应当遵循的5项关键指导原则：保障所有公民的使用权、保护公共利益、确保数据治理的核心地位、多利益相关方参与、以国际法为依据。

联合国贸易和发展会议（简称为贸发会议）作为联合国在消费者保护方面的牵头机构，致力于推广《联合国消费者保护准则》，贸发会议认为人工智能技术的发展必须以消费者保护为核心，确保技术进步惠及所有人，为此贸发会议正在呼吁建立更多的意识提升机制和强大的全球监管框架。贸发会议针对目前人工智能治理现状提出了多项建议，包括通过提高人工智能素养来增强消费者和政策制定者能力，支持各国制定国家人工智能战略，动员私营部门采用合乎道德的人工智能实践以保护消费者福祉，并致力于构建公平且负责任的人工智能全球框架。

联合国人权事务高级专员办事处（OHCHR）强调人工智能技术的应用必须尊重和保障人权，并要求相关从业者必须严格遵循8项主要原则，其包括管辖权、道德和法律依据、数据基础、责任与监督、控制、透明度和可解释性、数据主体权利和保障措施，旨在确保人工智能技术在规划、开发和实施时，不会侵犯个人隐私和其他基本人权。人权高专办提出在新技术部署前要对人工智能进行尽职调查和影响评估，从而防范与人权法不符或具有高人权风险的人工智能系统的应用。

联合国教育、科学及文化组织（简称“联合国教科文组织”）于2021年设立全球人工智能伦理与治理观察站，通过了《人工智能伦理问题建议书》，阐述了人工智能系统引发的就业率下降、“数字鸿沟”、教育失衡等伦理问题，并提出各成员国应采取适当措施，确保市场公平竞争和消费者权益得到保护。联合国教科文组织提出了10项原则，包括界定人工智能应用范围与进行风险评估、加强系统的安全与保障、保护隐私权与数据、多利益相关方参与、建立审计与问责制度、优化数据透明度和可解释性、确保人类进行监督与最终决策、坚持可持续发展目标、提升公众对人工智能的意识与素养和保证公平与非歧视，以确保人工智能的发展以人为本且符合伦理。

国际电信联盟作为通信技术领域的领导机构，自2017年以来一直在举办“人工智能造福人类”全球峰会，旨在利用人工智能潜在的积极影响，推动人工智能技术在医疗保健、能源消耗、气候变化和人工生产力等领域的应用，助力人工智能治理实现可持续发展目标。国际电信联盟已发布及正在开发的与人工智能相关的标准共计超过200项，并通过发起全球人工智能与数据共享倡议及7个面向所有人的预标准化焦点组，加强了在人工智

能标准化方面的工作，预标准化焦点涵盖人工智能与健康、自动驾驶、机器学习等多个领域。

3.2.2 区域组织与全球标准的规则衔接机制

区域组织通过规则输出促进全球治理范式融合，经济合作与发展组织（OECD）致力于制定人工智能治理的指导方针，在2019年制定了《经合组织人工智能原则》，并于2024年进行了修订。《经合组织人工智能原则》作为首个政府间人工智能标准，强调对人权和民主价值观的保护和人工智能的稳健性、安全性与可靠性，为二十国集团的人工智能原则奠定了基础。在对人权和民主价值观的保护方面，要求人工智能从业者在整个人工智能系统的生命周期内尊重法治、人权、民主和以人为本的价值观，解决由人工智能引发的信息滥用和虚假信息等问题。在稳健性、安全性和可靠性方面，规定人工智能系统在其整个生命周期内都应稳健、安全且可靠，建立适当机制确保人工智能在出现风险时能被安全干预、修复或退役。经合组织成立了全球人工智能伙伴关系（GPAI）和人工智能政策观察站，以促进其成员之间的人工智能治理的发展。

2024年，欧盟通过《人工智能法案》，要求成员国将统一的规则作为国家立法来规范人工智能，为人工智能技术在欧盟范围内的应用提供了全面的法律监管框架。该法案基于风险评估方法，将与人工智能相关的风险从低到高进行分类监管，明确了开发、评估和使用人工智能系统的严格规则和透明度义务。《人工智能法案》明确提及了对消费者保护的严格要求，并强调多利益相关方的参与和包容性。同时，欧盟要求欧洲标准化委员会（CEN）和欧洲电工标准化委员会（CENELEC）制定配套的欧洲标准，以支持《人工智能法案》的实施。

2024年2月，东南亚国家联盟在《东盟人工智能治理与伦理指南》中提出透明度和可解释性、公平性和公正性、安全性、以人为本、隐私和数据治理、责任和诚信以及强健性和可靠性7项原则，为该

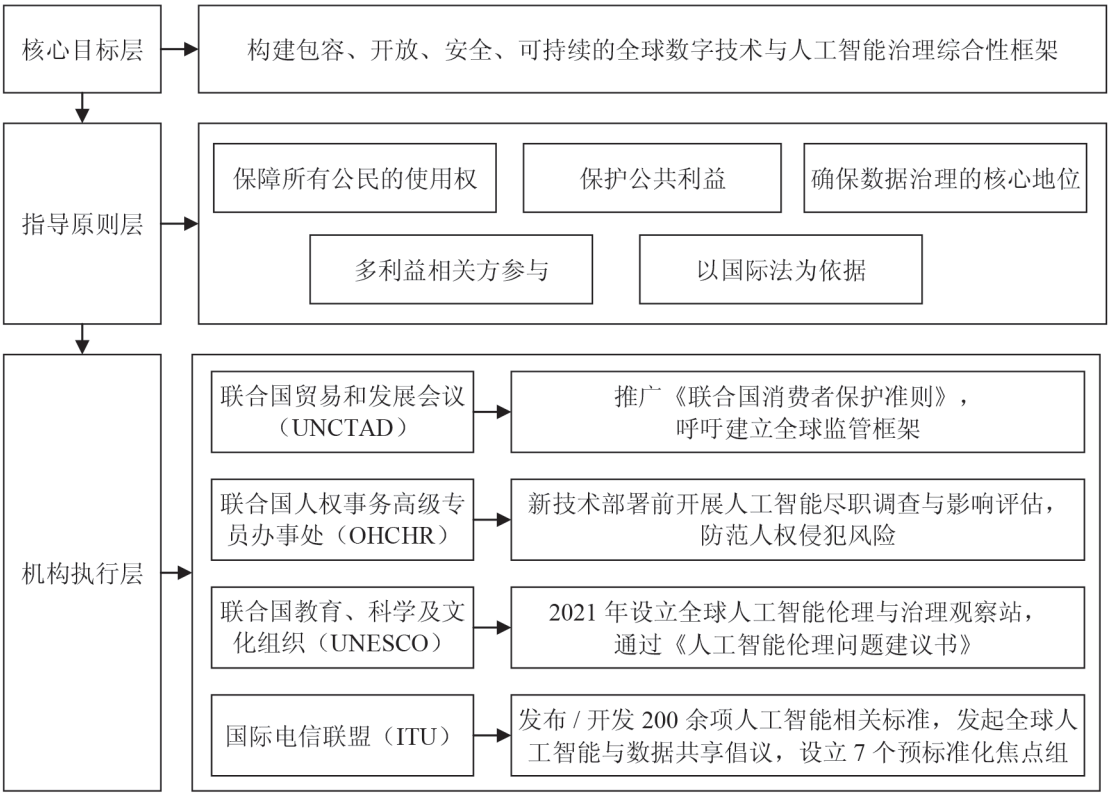

图1 联合国系统内跨机构治理框架

地区政府和企业在设计、开发和使用人工智能系统方面提供了非约束性指导，促进东盟成员国在人工智能领域的合作与协调。此外，东盟计划成立人工智能治理工作组，进一步推动区域共享标准的制定，加强对人工智能技术的规范和管理，以实现区域内人工智能技术的健康发展，提升地区整体竞争力。

2023年底，南方共同市场发布了《关于数字环境中民主和信息完整性的主席声明》，明确认识到人工智能滥用带来的威胁，如虚假信息传播、个人数据隐私侵犯等。声明呼吁成员国应实施适当的国家立法来保护隐私和个人数据，并鼓励科技公司采取透明、负责且尊重人权的政策，特别是在内容审核、推荐算法和个人数据处理方面，科技公司需要减少虚假或非法数据的传播，维护数字环境的健康和安全，保障消费者在数字时代的合法权益。

3.2.3 多元主体参与的标准共治机制

国际消费者协会（CI）作为全球消费者组织的代表，将人工智能治理作为消费者保护问题的优先事项，主要聚焦于四大核心领域，以保护消费者的合法权益。（1）要推动数字市场变革，加强数据隐私保护和信息透明化，保障基础网络接入的可负担性[20]；（2）确立高水平消费者保护标准，对人工智能系统的可信度和透明度进行独立监督；（3）构建包容性治理框架，促进多利益相关方参与政策制定；（4）完善救济与申诉渠道，确保消费者在算法歧视、数据滥用等问题中获得切实救济。

欧洲标准化消费者代表协调会（ANEC）作为欧洲消费者在标准化中的声音，在人工智能治理方面发挥着重要作用，提出国际标准应专注于技术要求，如算法透明度指标、数据安全技术规范等以促进技术稳健性、安全性和互操作性，但不应涉足公共政策领域，要将法律法规、伦理原则的制定交由政府立法与国际公约处理。ANEC认为在技术标准制定中应纳入消费者、劳工、环境组织代表等利益相关方，从而使标准能够平衡各方利益。此外，ANEC强调人工智能监管和风险治理领域应由政府主导，不能交由私营企业或自愿性自我监管，从而保证人工智能监管的公正性与强制性，避免企业因利益原因忽视对人工智能的监管，利用人工智能技术损害消费者利益。

4　消费者权益视角下的人工智能治理关键标准化路径

4.1 依托国际规范与多元共识构建标准基准

协同治理理论强调多元主体通过协商协作参与公共事务治理，以平衡各方利益并提升治理效能。基于这一理论，标准制定需系统整合国际规范与多方诉求：一方面吸纳联合国《人工智能伦理问题建议书》、经合组织修订版《人工智能原则》、欧盟《人工智能法案》等框架中关于数据保护、算法公平的核心要求，将其转化为可操作的技术标准；另一方面通过建立跨司法管辖区协商机制，纳入消费者组织、弱势群体及企业代表的实质性建议，使标准既体现技术可行性，又反映包括消费者在内的多元利益诉求，最终形成具有全球适用性与公信力的共识性规范，旨在通过多元主体参与破解监管碎片化难题，为消费者权益提供统一的标准保障。

4.2 建立包容性风险治理与系统设计范式

风险治理理论主张通过“风险识别—评估—应对—监控”的全周期管理实现风险可控。以此为支撑，人工智能治理框架需构建全周期风险管理体系：在风险识别环节，通过人权尽职调查与算法影响评估，精准定位算法偏见、数字鸿沟等可能损害消费者权益的风险点；在风险应对环节，强制贯彻包容性设计原则，如通过反歧视框架消除算法偏见以保障消费者公平交易权与尊严权，在应用层保留人工干预选项与替代性服务通道以维护自主选择权；在风险监控环节，将基础设施适配性与技能普惠要求纳入技术标准，持续遏制技术应用的结构性不平等，通过风险治理理论的系统性方法，确保消费者权益在AI系统设计、开发与应用的全流程中得到动态保障。

4.3 强化以消费者为核心的全周期权益保障

消费者主权理论强调消费者在市场交易中享

有主导地位，其知情权、选择权、数据主权等应得到充分尊重与保障。基于这一理论，需在AI研发至商业化全流程中构建权益保障体系：通过明确数据收集范围、存储期限及共享边界以保障数据主权，强制加密与全链路可追溯以维护隐私权与安全保障权，落实消费者的知情权；通过确立“人为控制”刚性标准（高风险系统设人工审核节点）、公共服务领域保留非AI服务选项，确保消费者对服务方式的自主选择权与平等获取权；通过畅通救济渠道，强化权益受损后的补救机制。将消费者从被动接受者转化为权益主导者，实现技术应用与权益保障的深度绑定[21]。

4.4 完善标准合规的协同实施机制

协同治理理论强调多元主体通过权责划分与功能协同提升治理效能，据此构建“认证—监管”双轨执行体系。在主体协同层面，推动标准制定组织与监管机构协作（政府主导监管）、引入第三方认证机构（独立评估）、设计企业激励政策（市场主体自律），形成“政府—市场—社会”的多元共治格局；在机制落地层面，通过合规审计工具开发、认证计划实施，确保标准条款转化为企业可操作的合规要求，并将核心标准纳入国家监管框架，实现技术规范与法律责任的闭环衔接。通过多元主体的功能互补与权责衔接，避免单一治理主体的局限性，最终通过标准的有效执行保障消费者权益落地[22]。

5 结 论

人工智能的全球化发展在重塑消费生态的同时，也因技术迭代与治理滞后的矛盾、监管碎片化风险及消费者权益保障机制的弱化，对全球治理体系构成严峻挑战。本文从消费者权益保障视角出发，通过双重效应分析揭示：AI技术虽通过普惠服务升级与效率赋能，显著提升了消费者的知情权、自主选择权与公平交易权，但数据滥用、算法偏见、技术垄断等风险，也直接加剧了对消费者隐私权、安全保障权、人格尊严权的侵害，凸显了技术红利与权益受损的现实张力。各国治理实践的差异表明，单一国家监管难以应对跨国消费场景下的权益保障困境，亟须构建以消费者权益为核心的标准化治理体系。

基于此，全球人工智能治理需依托四维协同机制筑牢消费者权益防线：（1）整合国际规范共识，将联合国、经合组织、欧盟等框架中涉及消费者数据保护、算法公平的要求转化为可操作的技术标准，确立消费者在数据与算法层面的权益基准；（2）建立包容性设计范式，通过全周期风险管理和反歧视框架，消除算法偏见对特定消费群体的排斥，保障每一位消费者平等享有AI技术的权益；（3）强化全周期权益保障，以数据主权与知情权为基础，通过设置人工干预节点、保留非AI替代服务等措施，确保消费者对AI服务的自主选择权与获取权；（4）完善标准合规协同机制，通过“认证—监管”双轨制推动标准落地，形成“技术规范—法律责任—权益救济”的闭环，让消费者在权益受损时能获得切实保障。将消费者权益保障嵌入标准设计与实施的全流程，通过多元主体共治与区域规则衔接，实现技术创新与消费公平的协同发展，最终构建以消费者权益为价值导向的安全、可信、普惠的全球人工智能治理体系。

参考文献

[1] 薛澜,赵静.人工智能国际治理：基于技术特性与议题属性的分析[J].国际经济评论,2024(3):52-69.

[2] WEF. “Global Risks Report 2024” [Z].

[3] 高秦伟.人工智能标准规制的监督机制[J].郑州大学学报(哲学社会科学版) ,2025,58(3):44-52.

[4] 敦帅,陈强,贾婷.中国人工智能治理研究述评与展望[J].中国科技论坛,2025(4):31-42.

[5] 江红艳,王海忠.人工智能时代下的消费者权益保护[J].人民论坛,2023(5):66-69.

[6] HERMANN E,PUNTONI S.Artificial intelligence and

consumer behavior:from predictive to generative AI[J]. Journal of Business Research,2024,180:114720.

[7] 李智迅,段军山.金融消费者权益保护：挑战、实践和路径[J].消费经济,2025,41(3):89–100.

[8] 潘青.我国互联网金融消费者权益保护体系建设与优化路径[J].经济体制改革,2020(1):196–200.

[9] 李鑫.大语言模型训练阶段的隐私风险及应对策略[J/OL].情报杂志,1–11[2025–07–25].http://kns.cnki.net/kcms/detail/61.1167.G3.20250703.1118.006.html.

[10] 毕德旭,常丽萍.基于人工智能的医疗数据分析中患者隐私权的保护机制研究[J].中国医学伦理学,2025,38(9):1184–1190.

[11] 徐实.平台用户知识隐私的法律保护路径：以DeepSeek式人工智能为中心[J].暨南学报(哲学社会科学版),2025,47(4):45–66.

[12] AI Standards Hub. Standards Database[EB/OL]. (2025–04 –14) [2025–04–28].http://www. aistandardshub. org / ai —standards — search.html.

[13] 陈剑锋.大语言模型在临床医学的可应用性探讨[J].医学与哲学, 2023,44(21)：1–6.

[14] 翟运开,罗波,王宇,等.患者医疗数据共享意愿影响因素：结合改进计划行为理论（TPB）与技术接受模型（TAM）的分析[J].科技管理研究, 2023,43(16):235–244.

[15] 江红艳,王海忠.人工智能时代下的消费者权益保护[J].人民论坛,2023(5):66–69.

[16] 刘晓.我国大数据征信个人敏感数据保护困境及保护机制研究[J].西南金融,2019(1):30–36.

[17] 胡弘弘,王惠民.人工智能治理的反身型转向：主要风险、全球模式与中国进路[J].图书馆建设, 2025(3):25–35.

[18] 王天禅.美欧人工智能治理的分化：基于治理结构、能力势差和战略选择的考察[J].国际关系研究, 2025(2):107–135.

[19] 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.

[20] 黄景贵,刘响俊,李东敖.国家大数据综合试验区对数字经济的影响研究:基于双重差分的实证分析[J].成都理工大学学报(社会科学版),2024,32(1):71–90.

[21] 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.

[22] 卫德佳,李珊.乡村振兴战略下财政衔接资金的法律监管研究[J].成都理工大学学报(社会科学版),2024,32(3):12–21.