世界互联网大会智库合作计划系列成果

全球人工智能治理的立法观察：经验与展望

武汉大学竞争法与竞争政策研究中心
新疆大学法学院
武汉大学网络治理研究院

2026年4月


世界互联网大会
World Internet
Conference


武漢大學
WUHAN UNIWERSITY
新疆大學
XINJIANG UNIVERSITY

研究团队

孙晋

武汉大学法学院二级教授

武汉大学竞争法与竞争政策研究中心主任

武汉大学网络治理研究院执行院长

帕孜丽娅·玉苏甫

新疆大学法学院讲师

前言

人工智能技术的飞速发展在带来创新机遇的同时，也引发了数据安全、算法偏见和责任认定等复杂治理挑战。为了系统梳理全球人工智能治理的立法现状并探索未来方向，报告深入分析了美国、欧盟、中国等主要国家和地区的立法实践，比较了不同治理模式的特点与差异。报告采用案例分析和比较研究的方法，重点考察了各国在数据治理、算法监管和损害责任等核心领域的制度设计。研究发现，各国立法模式呈现分化：美国侧重创新友好的弹性监管，欧盟建立了基于风险分级的严格框架，而中国则探索发展与安全并重的本土路径。目前全球治理面临技术迭代过快导致法律滞后、国家间监管标准差异阻碍协作以及数字鸿沟加剧社会不公等突出瓶颈。未来立法需要增强敏捷性，例如引入监管沙盒机制；深化国际合作以弥合规则分歧；并更关注技术普惠，确保人工智能发展符合人类共同利益。研究认为，平衡技术创新与风险防控，构建包容、协作的全球治理体系，是人工智能健康发展的关键。

目录

一、全球人工智能发展与治理立法的背景

(一)技术演进与产业格局：人工智能的发展态势

1.核心技术突破与主要类型分布

机器学习作为人工智能的基石，在监督与非监督学习的框架下，正持续突破复杂模式的识别能力。[1]作为其重要分支，深度学习依托多层神经网络，显著提升了图像识别(如医疗影像分析)和自然语言处理精度；循环神经网络(RNN)则进一步优化了时序数据的建模能力。[2]Transformer 架构的出现推动了自然语言处理的质变，大模型借助自注意力机制精准捕捉深层语义关联，为生成式人工智能(Generative AI)的爆发奠定了基础。[3]

从发展路径来看，人工智能主要呈现“专用型”与“通用型”双轨并进的态势。专用人工智能(Narrow AI)高度聚焦特定领域的垂直任务，如医疗诊断系统[4]和金融风控模型，[5]其运行高效，但存在场景局限与跨领域迁移壁垒。[6]相比之下，通用人工智能(AGI)雏形通过大规模预训练模型模拟跨领域智能，展现出强大的多任务泛化潜力，[7]但面临推理能力不足与伦理风险。

专用 AI 因成熟度高，已在工业自动化等领域规模化应用。通用 AI 虽处初期，却带动云计算等产业链扩张，催生新兴业态。技术多元化促使产业生态分层：底层算力芯片、中层算法框架、应用层行业方案与消费产品。此结构加速创新扩散，亦引发标准滞后与监管挑战。[8]

2.产业规模扩张与全球空间布局

全球人工智能产业规模持续扩大，2023 年市场总规模已突破数千亿美元，[9]增长源于算法优化、算力提升及在医疗诊断、金融风控、智能制造和智能交通等领域的深度应用。北美地区凭借强大研发实力和成熟资本占据核心地位，引领基础算法、高性能芯片及云服务创新。欧洲则更侧重特定领域应用深化和伦理框架构建，在工业自动化、医疗健康 AI 及数据隐私保护等方面形成特色，[10]但成员国间技术转化与产业化速度存在差异。[11]

亚洲，尤其是东亚，成为增长最快区域。中国、日本、韩国在政府战略与市场需求推动下产业生态迅速成型。[12]中国在计算机视觉、自然语言处理等应用技术及电子商务、智慧城市落地方面突出。但亚洲在基础理论与核心硬件研发上相较北美仍有追赶空间。整体而言，各区域在研发侧重、比较优势与治理进程上的分化，使得全球人工智能产业呈现出显著的非对称发展格局。

3.未来演进规律与关键发展趋势

人工智能正呈现多维融合态势，量子计算赋能机器学习在医药研发等领域实现底层突破，AI 与生物技术的深度耦合重塑了生命科学范式。应用场景由单点向元宇宙及智慧城市等全域跨越，依托实时数据驱动系统级智能跃升，倒逼算力基础设施向

[1] 薛澜, 王净宇. 人工智能发展的前沿趋势、治理挑战与应对策略[J]. 行政管理改革, 2024, 15(8): 4-13.
[2] 陈翎翔. 人工智能医疗器械大数据清洗的法律规制路径[J]. 东北大学学报 (社会科学版), 2025, 27(5): 117.
[3] 谢潇, 罗世杰. 论生成式人工智能的动态风险及适应性治理[J]. 北京工业大学学报 (社会科学版), 2025, 25(1): 112-125.
[4] World Health Organization. Ethics and governance of artificial intelligence for health: large multi-modal models. WHO guidance. World Health Organization, 2024.
[5] Pantanowitz, Liron, et al. "Regulatory aspects of artificial intelligence and machine learning." *Modern Pathology* 37.12 (2024): 100609.
[6] 付新华. 全球人工智能立法的多元趋势与中国模式[J]. 交大法学, 2025 (6): 60-73.
[7] 周洪宇, 李宇阳. 生成式人工智能技术 ChatGPT 与教育治理现代化——兼论数字化时代的教育治理转型[J]. 华东师范大学学报 (教育科学版), 2023, 41(7): 36.
[8] 季卫东, 赵泽睿. 法律如何实现动态治理——人工智能不确定性的立法应对[J]. 学术月刊, 2025, 57(3): 98-109.
[9] 叶淑兰, 李孟婷. 全球人工智能治理: 进展, 困境与前景[J]. 国际问题研究, 2024, 66(5): 100-118.
[10] Walter, Yoshija. “Managing the race to the moon: Global policy and governance in Artificial Intelligence regulation—A contemporary overview and an analysis of socioeconomic consequences.” *Discover Artificial Intelligence* 4.1 (2024): 14.
[11] 张辛鑫, 冀瑜. 论欧盟《人工智能法案》中的标准化模式及对我国启示[J]. 标准科学, 2024, 5: 39-46.
[12] 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.

云边协同重构。在产业生态层面，巨头并购加速技术内化，开源社区促进技术民主化，全球监管差异引发产业布局调整，生成式 AI 则触发了出版等行业价值链的深度重组。

(二)风险溯源与法律挑战：人工智能引发的现实困境

1.数据侵权风险与个人隐私保护

人工智能应用面临多样化的数据安全风险，其核心诱因在于数据滥用、非法采集与隐私泄露。数据滥用多表现为超授权使用，如将健康生理数据擅用于商业分析；非法采集则聚焦于后台过度抓取位置及通讯录等敏感信息；隐私泄露则涉及数据库受损或算法对个体特征的逆向还原。此类风险不仅直接侵害个人权益，更易诱发深层的歧视性后果。

现行法律应对存在不足：数据确权不清导致用户维权困难；使用边界模糊，二次利用常超原始同意范围；侵权救济受阻，受制于高昂的举证门槛及平台以“技术中立”为由的避责抗辩。[13]

治理需构建多层框架：欧盟 GDPR 的严格规则虽具参考价值，但可能抑制创新。更可行的是实施风险分级监管，严控医疗等高危场景，放宽基础研究数据流动。[14]同时强化技术工具，如差分隐私和可验证删除机制。[15]中国算法备案与安全评估制度提供了动态监管样本。最终目标是在保障核心权利的前提下，为负责任 AI 发展保留空间，需立法持续调适规则。

2.算法黑箱效应与公平正义博弈

AI 算法决策的透明性与可解释性缺陷引发社会公平争议，黑箱效应导致偏见放大，如招聘性别歧视和金融拒贷案例。行业治理路径各异：医疗领域强制算法解释报告，保留医生否决权；电商系统采用用户标签自主管理；欧盟《人工智能法案》要求高风险场景技术文档记录。提升算法可审计性与问责是关键：技术层面发展反事实解释工具，法律层面建立分层责任机制，如医疗 AI 误诊时开发者证明合规、用户承担修改责任，并建立第三方认证机构进行偏见检测。

3.侵权责任归属与法律主体认定

人工智能系统侵权责任认定面临法律困境，传统归责框架难以适配其技术特性。争议焦点集中于产品责任、使用责任与开发责任。产品责任视角下，制造商对视为普通商品的 AI 系统应承担严格责任；但系统自主学习常使损害脱离设计预期，导致因果关系复杂化。使用责任强调操作者过失，但 AI 高度自主性削弱人类直接控制力，如 L3+自动驾驶中驾驶员仅为“监督者”，注意义务边界模糊。开发责任涉及算法设计缺陷认定，深度学习“黑箱”特性使追溯代码异常困难，医疗 AI 误诊常因无法验证决策而陷入责任真空。

典型案例凸显司法困境。德国 2022 年自动驾驶致死案依《产品责任指令》判制造商主责，但暴露多主体协作系统责任划分的结构性缺陷。IBM Watson 肿瘤系统误诊纠纷中，医院与开发商对算法责任归属争议，揭示“人机协同决策”的规制空白。

解决路径需构建分层责任机制。具物理载体的智能产品(如工业机器人)可强化产品责任，要求制造商建立算法决策日志系统。生成式 AI 等软件系统适用“开发-部署-使用”三级框架：开发者担基础模型安全义务，部署者行场景测试责任，使用者证操作合规。[16]借鉴欧盟《人工智能法案》高风险系统强制责任保险，实现风险社会化分散。根本突破在于立法确认 AI“有限法律人格”，赋予其承

[13] Fernández, José Vida. "Artificial intelligence in government: Risks and challenges of algorithmic governance in the administrative state." *Ind. J. Global Legal Stud.* 30 (2023): 65.
[14] Ebers, Martin. "Truly risk-based regulation of artificial intelligence how to implement the EU’s AI Act." *European Journal of Risk Regulation* 16.2 (2025): 684-703.
[15] Lami, Bareq, et al. "The role of artificial intelligence (AI) in shaping data privacy." *International Journal of Law and Management* 68.2 (2026): 296-318.

担程序性责任能力，同时保留人类主体最终责任归属。[17]

(三)治理诉求与立法动因：规范人工智能的必要性

1.防范技术异化风险的现实需要

防范技术异化风险是人工智能治理的紧迫议题。技术系统可能脱离人类预设目标，展现不可预测行为，形成自主性失控。例如在金融或军事系统中，算法自我迭代可能超出设计边界，偏离人类价值基准。其根源在于深度学习模型的高度复杂性与黑箱特性，威胁系统安全并侵蚀人类认知主体性。

人工智能深度介入知识生产与社会决策时，过度依赖算法可能导致个体思维同质化，削弱批判性思考；在医疗等领域，依赖 AI 辅助或致专业技能钝化。这反映出技术工具反噬人类认知能力的风险，本质是人机关系异化，技术反向塑造并限制人类认知维度。

面对此类深层次累积性风险，传统事后追责机制存在局限。需建立预防性法律机制，在技术研发初期介入引导。通过设置算法透明度基线、强制部署人类监督回路、建立关键系统熔断机制等制度设计，将伦理安全要求内化为技术规范，构建制度性护栏，确保人工智能发展始终服务于人类福祉。

2.维护公众基本权益的安全底线

人工智能技术的快速发展对公民隐私权、公平权等基本权利构成新挑战。数据聚合分析解构传统隐私概念，生物特征识别加剧个人信息暴露风险。算法公平性问题突出，招聘、信贷等领域的自动化决策系统可能基于历史偏见对特定群体形成系统性歧视，且其隐蔽性导致救济困难。

不同司法管辖区采取差异化保护路径：欧盟通过 GDPR 设定严格的事前合规义务；美国依赖事后司法追责，应对大规模侵权效率不足；中国《生成式人工智能服务管理暂行办法》建立分级分类治理框架。这些模式均面临技术迭代超越法律更新的结构性矛盾。

构建有效保障体系需突破传统立法思维：在技术研发初期设立伦理边界，如禁止深度伪造侵害肖像权，或在医疗领域保留人类决策权；建立动态算法评估机制，通过第三方审计持续监测权利风险。该设计既保留创新空间，又以负面清单划定刑事司法、社会保障等关键领域的权利红线。

3.引导产业健康有序发展的制度保障

人工智能产业迅猛发展伴随市场失灵，威胁长期健康发展。突出表现为垄断风险：技术研发需庞大数据、顶尖人才及巨额算力投入，天然高门槛叠加网络效应，易致资源向头部企业过度集中，形成“赢家通吃”，削弱竞争、挤压中小企业并抑制创新。另一关键问题是复杂化的不正当竞争，包括数据垄断、算法合谋、市场封锁及歧视性定价，更隐蔽的如深度伪造虚假宣传、“大数据杀熟”等新型手段，扭曲市场信号、损害消费者权益。

应对关键在于产业政策与法律规制的深度协同。产业政策可引导方向、培育生态，如设立基金支持基础研究、建设开放数据平台。但仅靠鼓励性政策不足以解决垄断与不正当竞争，甚至可能加剧市场集中。必须通过强力法律规制构建清晰市场规则，提供公平竞争环境。核心是建立适应人工智能特点的竞争规则体系：明确数据垄断、算法协同等新型垄断的审查标准；细化算法歧视、“杀熟”等不正当竞争行为的禁止与处罚；健全数据确权、流通与公平获取的法律框架。

法律规制需具前瞻性，适应技术迭代与产业演进，在鼓励创新、激发活力与防范市场扭曲、维护公平间取得动态平衡。这要求深刻理解技术逻辑及对市场结构的影响，设计兼具原则性与操作性的规则，辅以有效监管执法，引导产业健康有序发展。

[16] Laux, Johann, Sandra Wachter, and Brent Mittelstadt. "Three pathways for standardisation and ethical disclosure by default under the European Union Artificial Intelligence Act." *Computer Law & Security Review* 53 (2024): 105957.
[17] Ruschemeier, Hannah. "AI as a challenge for legal regulation–the scope of application of the artificial intelligence act proposal." *Era Forum.* Vol. 23. No. 3. Berlin/Heidelberg: Springer Berlin Heidelberg, 2023.

二、主要国家和地区人工智能治理的立法实践

(一)美国：创新驱动下的弹性立法模式

1.顶层设计：国家战略与软性规制框架

美国人工智能治理采用软性规制模式，通过非强制性政策框架引导行业，注重规制的灵活性。2019 年《人工智能倡议》确立技术领先与创新生态目标，依托行业自律准则、技术标准(如 NIST 无法律约束力的风险管理框架)及跨部门协调机制。政策演进呈渐进特征，以行政命令、白皮书和国会听证为主渠道，2020 年《人工智能应用监管指南备忘录》要求联邦机构进行 AI 影响评估，为不同场景留出弹性空间，体现技术创新与风险防控的动态平衡。

该框架为市场主体提供了多元的发展空间，支持科技巨头参与伦理共建，并助力初创企业利用监管沙盒开展测试。其核心逻辑在于激活市场调节韧性，依托产业自律与政策灵敏度的深度耦合。目前，联邦政府正探索在高风险领域引入更具确定性的指引，旨在进一步优化软性治理的制度边界，确保治理效能与技术演进同频。

2.领域深耕：关键应用场景的专项立法

美国在自动驾驶立法中实行阶梯式监管，允许低风险试点并要求 L4 及以上车辆配置数据记录设备；医疗 AI 领域延续风险分级，诊断算法需经临床验证，而辅助研发工具则侧重算法偏见审查。数据治理同样采取场景化策略：自动驾驶侧重路况数据脱敏，医疗领域强调患者生物特征保护。此种纵向立法构建了“行业-风险-数据”的三维矩阵，例如自动驾驶事故中引入的“技术缺陷追溯期”制度。有别于欧盟的横向通用框架，美国的垂直领域规则更注重场景针对性，能提供具体的合规指引(如医疗算法需具备临床可解释性)，未来针对跨领域技术融合的规则协同与衔接将成为其演进方向。

3.成效评估：立法实施的产业影响分析

美国 AI 立法模式呈现出显著的阶段性效应。其弹性框架初期有效助推了产业创新：相对宽松的环境促使 2020 至 2023 年间风险投资激增近三倍，自然语言处理与计算机视觉领域涌现出众多突破性企业。这种轻触式的治理路径降低了企业的合规成本，加速了技术转化，例如自动驾驶企业在道路测试阶段的算法专利实现了逾 25%的年增长。

随着技术演进，相关制度设计面临新的适应性挑战。基础模型领域的规则尚待完善，大语言模型训练数据的合规边界亟待厘清，这客观上引发了 2022 至 2023 年间多起数据权益纠纷。在系统性风险防范机制的探索中，算法的不透明性在金融、医疗等场景持续引发广泛关注，相关公平性申诉五年内增长七倍。同时，当前的规则环境也在一定程度上推高了基础模型研发的市场集中度，初创企业的市场份额暂不足 15%。[18]

现行机制在统筹技术创新与风险管控时面临协调难题。《算法问责法案》等分散式立法实践导致州际标准存在差异：加州倾向于要求算法年度审查，而德州则允许企业以商业秘密为由保留关键参数。这种分散的治理模式客观上增加了企业的跨州运营成本，也在一定程度上分散了整体的风控效能。面对生成式 AI 的爆发，现有规则体系在深度伪造治理及侵权认定方面面临适用性考验，超 60%的相关案件因具体裁判标准尚待确立而中止调查。

(二)欧盟：权利基石上的综合立法模式

[18] Birkstedt, Teemu, et al. "AI governance: themes, knowledge gaps and future agendas." *Internet Research* 33.7 (2023): 133-167.

1.制度衔接：GDPR 对人工智能的先行规制

GDPR 的核心原则深刻影响着 AI 规制，目的限制与最小必要原则要求 AI 的数据处理必须用途明确且范围适度。在医疗诊断 AI 开发中，训练数据须与诊疗目标直接关联，超范围收集基因数据即构成合规风险。该条例第 22 条赋予了数据主体拒绝纯算法决策的权利，从而促使企业引入人工复核机制。例如，荷兰市政厅曾因算法未能解释福利发放的决策逻辑而承担法律责任，被认定未能充分保障数据主体的知情权。该条款在实质上确立了算法可解释性的初级标准。

生成式 AI 的海量训练模式客观上对传统的“目的限制”原则提出了新的适用性课题。尽管利用公开数据进行模型训练具备一定的合规基础，但其输出端仍存在重构个人信息的可能。在此情形下，由于原始数据已被转化为难以直接溯源的参数权重，数据主体往往难以有效行使删除权。这一现象折射出现行规则在数据衍生价值治理上尚有探讨空间，也反映出传统数据保护框架在回应新型技术范式时，仍需在制度供给层面作进一步的拓展。

2.核心解读：《人工智能法案》的风险分级机制

欧盟《人工智能法案》构建了四级风险分类体系，呈现上窄下宽的金字塔结构。顶层“不可接受风险”类禁止开发操纵人类潜意识的系统或政府社会信用评分工具。其下的“高风险”领域对医疗设备、关键基础设施等场景实施严格监管，要求提交涵盖算法逻辑、训练数据及误差评估的技术文档，并确保人类可随时介入控制。该层级同时要求训练数据应当防范偏见，相关技术档案须接受监管部门审查。[19]

“有限风险”与“极低风险”层级(如对话机器人)主要被赋予使用告知义务。[20]此种分级机制依据应用后果而非技术门类设定监管强度，未来在生成式 AI 等新兴技术场景下，其分类边界的适用规则仍有待实践进一步精细化。在兼顾中小企业合规资源投入的同时，该法案实质性地将抽象风险转化为可量化的管控指标，为全球人工智能治理提供了具有参考价值的制度范本。

3.溢出效应：布鲁塞尔效应下的全球立法示范

欧盟《人工智能法案》通过“布鲁塞尔效应”影响全球人工智能治理，促使部分非欧盟国家为降低跨国合规成本或提升规则兼容性，选择参照其技术规范。以巴西为例，其监管草案在“高风险系统”的界定及强制性义务上，深度借鉴了欧盟的风险分级理念；加拿大《自动化决策指令》则吸收了算法透明度与人工监督要求，以强化公共部门的 AI 问责效能。此种规则扩散，很大程度上依托于欧盟单一市场的大市场效应及其相对成型的监管体系所具备的先例价值。

在全球化进程中，欧盟规则与其他治理路径呈现出多元互动的态势。美国偏向于行业自律与垂直领域治理，依托联邦贸易委员会的个案执法及特定领域的专项规则，与欧盟的综合性立法展现出不同的治理逻辑。中国路径则注重统筹发展与安全，如对生成式 AI 实施包容审慎的备案制管理，为本土技术创新预留空间，在相关数据治理等议题上与欧盟标准呈现出各具特色的制度偏好。此外，欧盟规则相对详备的合规要求(如针对高风险系统的技术文档与持续监测机制)，客观上对企业的合规资源配置提出了更高标准，这也是其在跨国适用中需持续探索优化的方向。因此，欧盟模式在未来的全球治理体系构建中，将持续与各国的制度环境及技术

[19] Van Kolfschooten, Hannah, and Janneke Van Oirschot. "The EU artificial intelligence act (2024): implications for healthcare." *Health Policy* 149 (2024): 105152.
[20] Neuwirth, Rostam J. "Prohibited artificial intelligence practices in the proposed EU artificial intelligence act (AIA)." *Computer Law & Security Review* 48 (2023): 105798.

演进深度交融，其示范效应的广度与深度仍将由实践进一步塑造。

(三)中国：分类分级的发展性立法探索

1.政策引领：从产业促进到规范治理的变迁

中国人工智能政策演进呈现出显著的阶段性特征。2017 年出台的《新一代人工智能发展规划》确立了国家战略，聚焦技术突破与产业拓展，依托资源引导与平台建设赋能行业，初期治理更侧重于释放创新活力。随着生成式 AI 等前沿技术的深度应用，数据合规与算法公平等治理命题日益受到重视，政策重心有序向规范治理延伸。从《互联网信息服务算法推荐管理规定》到《生成式人工智能服务管理暂行办法》的相继落地，清晰映射出政策导向向统筹高水平安全与高质量发展的稳步转型。

“包容审慎”的治理理念在实践中得到了多维度的贯彻：在包容性方面，注重为新兴业态预留必要的探索空间，例如针对生成式 AI 引入备案机制以优化合规流程；同时，分类分级监管有效落实了差异化规制，切实保护了市场创新动能。在审慎性方面，则聚焦于关键风险点的精准管控，通过细化安全评估、信息内容审核、数据来源审查以及深度合成内容标识等要求，筑牢了技术伦理与市场秩序底线。[21]

这一政策范式实现了向“发展与规范并重”的跨越，其内在逻辑是基于技术成熟度与风险认知的适应性调控。在发展初期充分释放产业动能并积累治理经验；当技术应用深度触及公共利益时，及时通过立法厘清行为边界。相较于欧盟偏向综合前置规制的路径以及美国的行业自律模式，中国方案更强调动态风险评估与渐进式的法律约束，致力于构建安全底线与创新活力相平衡的现代治理架构。

2.深度切入：算法、大模型及垂直行业的立法尝试

中国人工智能立法高度聚焦特定技术场景，算法推荐、深度合成及大模型已成为现阶段的重点规制领域。《互联网信息服务算法推荐管理规定》确立了用户权益保障框架，要求平台落实透明度机制并提供便捷的关闭选项，同时依据算法属性设定差异化的合规义务。《互联网信息服务深度合成管理规定》明确了生成内容的标识要求，贯通全流程管理链条，合理划分了服务提供者与技术支持者的权责边界，并引入常态化安全评估机制。针对大模型治理，当前以备案制为核心抓手，要求研发主体依法提交关键技术信息，监管部门则依托持续监督与安全审查，形成了“先备案、后监管”的动态治理闭环。在自动驾驶等垂直领域，主要依托地方立法先行先试，与国家层面的顶层设计形成良性互动，共同勾勒出具有中国特色的渐进式治理图景。

3.范式总结：平衡发展与安全的中国本土实践

中国在人工智能治理领域已逐步构建起“动态平衡”范式，有效统筹了技术创新与风险防控。该范式主要依托以下制度工具：

(1)安全评估：针对较高风险的 AI 系统，在部署前开展技术可靠性与社会影响审查，通过设定合理的参数阈值引导研发主体规避潜在风险；

(2)算法备案：要求提交算法逻辑等核心机制，相较于全景式的披露规则，此举更侧重于关键节点的透明度，有效回应了算法不透明带来的社会隐忧；

(3)内容标识：要求明确标注生成内容的来源，以保障社会公众的知情权。

其底层支撑在于分类分级的治理体系：即依据应用场景的风险等级实施差异化规制。例如，对医疗诊断 AI 适用更为审慎的评估机制，而对工业质检则倾斜于事后监测。此种精准施策切实保护了市场主体的创新活力。

[21] 李苑君.人工智能法律治理的域外实践与中国方案[J].中南民族大学学报(人文社会科学版),2025,45(12):190-204+212.

面向未来，国内技术标准的协同互认，以及在数据跨境流动、算法认证等议题上与国际规则的深度衔接，将是制度演进的重要方向。进一步完善标准层面的顶层设计，并持续参与全球治理对话，将有助于提升该治理体系的国际兼容性与跨国协同效能。

(四)其他：多元视角下的本土立法借鉴

1.英国：基于原则的敏捷监管模式

英国采取基于原则的敏捷 AI 监管模式，其核心依托《人工智能监管白皮书》所确立的五大原则：安全性、可靠性与稳健性、透明性与可解释性、公平性以及问责制与治理。该指导性框架为各行业提供了灵活的合规参照。有别于欧盟详尽的成文规则体系，此种模式能够有效回应技术快速迭代的诉求，保持制度的适应性。在实施路径上，政府侧重运用非强制性政策工具推动治理落地，如制定行业标准、发布最佳实践指南及倡导自律规范。[22] 同时，“数字监管合作计划”有效促进了监管机构、产业界与学术界的多方协同，共同探索契合技术演进的治理方案。该模式实质上提供了一种轻量化的治理路径，通过原则框架引导市场自我约束，并辅以动态指南应对新兴挑战，有助于优化立法资源的配置效率，在统筹技术创新与风险管控方面展现出独特的实践参考价值。

2.加拿大：注重公职系统算法问责的立法

加拿大《自动化决策指令》针对公共部门算法治理确立了系统性的规制框架。该指令要求联邦机构部署 AI 系统前强制实施算法影响评估(AIA)，全面审查技术特征及潜在的数据偏见风险，并依法公开评估结果。在救济机制方面，保障相关主体获取简明决策逻辑解释的权利，并设立独立的人工智能申诉专员处理异议，必要时可中止算法运行。[23]其责任认定引入“动态问责”机制：在开发阶段要求留存完整的算法日志，运营阶段持续监测决策偏差，一旦触及阈值即触发人工干预。此种贯穿全生命周期的治理模式，有效确保了公权力对最终决策的把控，为统筹数字政府效能与权利保障提供了实践样本。

3.日本：以非强制性指南为主导的治理路向

日本在人工智能治理中侧重采用非强制性指南模式，这与欧盟的成文法路径形成对照。其核心文件主要依托行业自律与社会共识，旨在统筹技术创新与规范演进。政策制定过程强调“产学官”协作机制，通过凝聚政府、学界与产业界的合力，以提升规则的专业度与行业认同。其“敏捷修订”机制有效缩短了规则更新周期，例如在应对生成式 AI 深度合成风险时，能够快速增补内容标识条款。[24] 在规范落实方面，该模式主要通过软法进行引导，未来在医疗 AI 等高风险场景中的协同治理效能仍有待进一步观察。对于处于产业探索期的经济体而言，此种模式提供了具备较高灵活性的政策选项，有助于在发展初期凝聚共识，并为后续的制度构建积累实践经验。

[22] Roberts, Huw, et al. "Artificial intelligence regulation in the United Kingdom: a path to good governance and global leadership?." *Internet Policy Review* 12.2 (2023): 1-31.
[23] Scassa, Teresa. "Administrative law and the governance of automated decision making: A critical look at Canada's directive on automated decision making." *UBCL Rev.* 54 (2021): 251.
[24] Kozuka, Souichirou. "A governance framework for the development and use of artificial intelligence: lessons from the comparison of Japanese and European initiatives." *Uniform Law Review* 24.2 (2019): 315-329.

三、全球人工智能治理立法的内容维度与模式

(一)立法核心：关键要素的制度定性

1.资源层：数据流动、确权与安全保障

人工智能数据治理的核心维度涵盖数据跨境流动、产权归属与安全保障。数据跨境流动呈现出主权监管与全球协作之间的内在张力，部分经济体基于安全与隐私考量实施数据本地化策略，这在客观上对跨国科研合作与商业协同提出了更高要求。实践中，各方正积极通过双边协议或区域信任框架探索规则衔接的平衡点。

数据产权归属面临着权属界限亟待厘清的现实课题，特别是在医疗、金融等高价值数据应用场景中，相关权益主张亟需法律规范。对此，分层确权模型提供了一种探讨路径：即原始数据提供者保留所有权，数据处理者获得使用权，数据聚合平台享有收益权。目前，该模型在权利边界的精细划分及侵权救济机制的构建上，仍有进一步深化的空间。

数据安全保障有赖于技术、制度与责任的三位一体协同。在技术层面，广泛应用端到端加密、分布式存储及动态访问控制；管理规程侧重于权限分级与操作留痕；应急响应机制则明确了监管通报及受影响个体的告知程序。

2.技术层：算法透明度与可解释性要求

算法治理的技术层规制主要聚焦透明度与可解释性。实现透明度的技术路径各有侧重：适度公开源代码有助于增进公众认知及公共决策的正当性，实践中需注重与商业秘密保护相协调；披露关键模型参数(如欧盟针对高风险系统的要求)能揭示变量权重，但复杂模型的认知门槛仍需通过专业解读加以弥合；此外，规范技术文档以记录设计逻辑与数据特征也是常态化手段。[25]可解释性的适用标准呈现出显著的场景化差异。医疗诊断、金融等高风险场景倾向于深度解释，要求明确判定依据与关键因子来源；常规推荐系统等低风险场景则侧重基本原理的告知。对于解释深度的量化指标及司法裁判尺度的统一，实务界仍在持续探索。算法审计是推动上述规则落地的保障。这要求审计机构具备计算机、法律伦理及行业专长的复合资质，且流程需贯穿从数据偏见检测至部署后监测的全生命周期(医疗等特定领域需附加临床验证)。同时，提供可追溯案例并建立申诉与人工复核机制亦是核心环节。随着生成式算法涌现性特征的日益凸显，促使审计标准与前沿技术保持同频，将是后续制度演进的重点考量方向。

3.后果层：损害责任的认定原则与分配机制

人工智能损害责任的法律界定主要围绕归责原则适用、责任主体认定及损害赔偿机制三个核心议题展开。

归责原则的适用是首要探讨方向。传统过错责任要求证明主观过失或故意，在自动驾驶传感器误判等算法自主决策场景下面临举证挑战。严格责任则立足于行为危险性(如高风险医疗 AI 的诊断偏差)，侧重结果归责。在司法实践中，两者边界的划定通常需结合应用场景的风险等级与技术可控性进行个案考量，欧盟《人工智能法案》的风险分类体系在此提供了一定的参照。

产业链多主体的参与使得责任界定更具复杂性。开发者、运营者与使用者的边界往往存在交叉：开发者可能因算法逻辑或训练数据偏见承担责任(如生成式 AI 涉嫌侵权)；运营者主要回应系统部署后的运行风险(如推荐算法的差异化定价)；使用者的责任多见于违规操作或技术滥用(如深度合成技术的违规使用)。在多方主体并发的情境下，

[25] 贾开,赵静,周可迪.算法全球治理：理论界定、议题框架与改革路径[J].中国行政管理,2022,(06):59-65.

责任份额的精细化划分规则仍有待完善，当前在自动驾驶等领域的实务裁判较多依赖于个案裁量。

损害赔偿机制的构建同样面临技术层面的考验：1. 损失测算具有高度复杂性，AI 损害兼具隐蔽性与长期性(如算法偏见引发的隐性歧视)，要求评估时全面考量直接经济损失、人身权益及衍生损害；2. 赔偿限额的设定需妥善统筹产业发展空间与法益保障，高风险 AI 系统可适度借鉴医疗器械领域的强制责任保险等制度；3. 面对多源数据融合或模型不透明带来的因果关系认定难点，实务界正积极探索举证责任缓和规则以及引入特别技术鉴定程序。

(二)范式选择：治理路径的差异化表征

1.集中统一：涵盖全领域的综合立法范式

综合立法范式旨在确立覆盖技术全链条与应用全场景的总体规则框架。欧盟《人工智能法案》构建的四级风险分类体系(不可接受风险、高风险、有限风险、最小风险)，针对医疗诊断等高风险领域设定了数据规范、文档留存及人工监督等强制性合规基线。其依托标准化的认定流程与集约化的监管设置，所形成的“金字塔式”结构旨在提升规则的执行效能。

中国的相关实践以《新一代人工智能发展规划》为依托，展现出政策与立法协同演进的特征。核心机制涵盖算法备案与安全评估等全生命周期环节，并引入分类目录管理以动态识别风险，如实施生成式人工智能分级备案制。此种模式致力于统筹化解监管交叉与规则适用空白：欧盟路径客观上优化了跨国企业的合规预期，而中国的算法备案机制则显著增强了规制的穿透性。

在实施层面，综合立法的稳定性与技术迭代的敏捷性之间存在着持续的调适空间。欧盟现有的风险分类标准在回应通用人工智能(GPAI)等前沿发展时，其规则颗粒度尚待精细化；中国的制度设计虽已预留动态空间，基础模型训练数据的合规边界亦有待在实务中进一步明晰。面对医疗 AI 临床验证、金融实时风控等垂直领域的差异化诉求，如何在维持规则稳定性的同时赋予适度的制度弹性，是深化该治理范式的核心考量。

表 1 人工智能综合立法范式对比表

主体	核心特征	风险管控方式	主要优势	现存挑战
欧盟《人工智能法案》	统一法律框架；四级风险分类体系	预设规则；高风险领域强制义务	提升执法效率；降低跨国合规成本	风险分类滞后；GPAI 规则模糊
中国综合治理体系	政策立法协同；全周期监管网络	分类目录管理；分级备案；动态调整	解决监管真空/重叠；提升穿透力	新兴问题需细化；难满足行业特殊需求
综合立法范式共性	追求体系化；统一规则	全领域覆盖；预设/动态结合	稳定市场预期；避免分散立法弊端	刚性弹性难平衡；技术迭代适应弱

2.分散渗透：基于特定行业的分散立法范式

分散渗透式立法模式侧重基于不同行业的风险特征实施差异化规制。美国在医疗健康领域将基于机器学习的诊断软件纳入传统医疗器械审批框架，注重动态性能监测与实时质量追溯，以统筹技术创新与患者权益保障；金融科技领域则要求核心算法引入“反事实公平性测试”及人工干预否决机制，旨在提升决策透明度，防范潜在的算法偏见及系统性金融风险。

日本《道路运输车辆法》修正案围绕自动驾驶确立了基于运行设计域的分层责任制度。以 L3 级

别为例，规则要求实时监测驾驶员状态，并在系统接管请求未获及时响应时触发紧急制动。与之配套的事故数据记录设备强制标准，为事故责任的客观界定提供了法定证据支撑。

此种模式的核心在于突出的专业适配性：其根植于特定行业的技术逻辑(涵盖医疗生物特征处理、金融实时风控、自动驾驶环境感知等)，能够精准锚定技术维度的合规基线。在实践运用中，该模式亦面临多维度的规则调适诉求：跨行业场景下的多重审查往往推高企业的综合合规成本(如健康可穿戴设备数据应用于保险精算时触发的交叉管辖)；产品分类边界的游移可能引发规则适用的规避行为；跨部门协同监管机制的深化与整合，则是提升整体治理效能的关键所在。

3.动态均衡：硬法与软法协同的包容治理

人工智能的混合治理旨在统筹硬性规则与柔性规范的相互调适。英国的实践依托“原则+指南”架构，锚定安全性保障、决策透明度及基本权利尊重三大核心维度。监管机构协同技术与产业界共建行业指南，细化算法标准与合规路径，在提供明确指引的同时预留了充分的制度弹性。

新加坡实施“测试-学习-调整”的渐进式循环：创新主体在预设边界内开展应用测试，监管部门同步汇聚数据以评估风险，进而优化后续规范要求。此闭环机制有效平抑了传统规则稳定性与技术敏捷演进之间的固有张力。

软法规范(如行业标准、伦理准则)凭借程序上的灵活性，能够敏捷回应前沿场景的规则诉求，提供先期行为指引与风险缓释。硬法规范则侧重于划定安全底线、明晰权责分配，输出制度层面的确定性与刚性约束。软法重在引导业态探索，硬法致力于夯实法治基座，两者在功能层面的交互与互补为包容性治理提供了实践框架(表 3)。[26]

表 2 人工智能混合治理实践对比
(英国 vs 新加坡)

治理架构	核心机制	主要特点	应对挑战
原则+指南	核心原则+行业指南；多方协同制定	保留创新空间；避刚性阻碍	技术快速迭代
测试-学习-调整循环	预设边界试点；动态反馈优化规则	持续学习；动态响应	技术不确定性

表 3 混合治理中软法与硬法功能对比

形式	优势	局限性	核心价值
行业标准、伦理准则等	灵活；更新快；填补规制真空	约束力有限(自愿遵守)	初期指引；风险缓释
法律条文	强制约束；明确责任；处罚措施	难快速适应新场景	确定性保障；秩序基石

(三)全球协作：跨境治理的沟通与衔接

1.国际组织在法律规则协调中的枢纽功能

国际组织在人工智能全球规制协同中发挥着实质性的枢纽效能。经济合作与发展组织(OECD)倡导的基于人本理念的《人工智能原则》，已转化为多边政策的重要参考基准。联合国层面侧重评估技术演进对可持续发展议程的深远影响，致力于依托多边对话机制统筹不同经济体的治理诉求。二十国集团(G20)则聚焦技术变革对全球宏观经济及就业

[26] Qian, Yuzhou, Keng L. Siau, and Fiona F. Nah. "Societal impacts of artificial intelligence: Ethical, legal, and governance issues." *Societal impacts* 3 (2024): 100040.

领域的结构性影响，持续推进成员国间的政策衔接与互动。

表 4 全球 AI 治理三大国际组织对照

组织	规则性质	核心输出	影响范围
OECD	非约束伦理原则	《AI 原则》	38 成员国；全球参考
联合国 AI 专家组	多边对话平台	可持续评估报告	193 成员国
G20	高层经济对话	部长声明/政策建议	20 大经济体

2.多边框架下立法合作的典型案例剖析

在全球人工智能治理的多边合作实践中，欧盟-美国贸易与技术委员会(TTC)的协作机制具有代表性。该机制下设专职工作组，重点推进技术标准互认与风险管理协同。双方依托定期对话，围绕算法透明度框架、高风险系统分类标准等核心议题展开技术协调。以生物识别技术的应用为例，双方共同研讨隐私保护基线，推动形成非约束性技术指南。此种协作模式为跨国企业提供了合规参照，客观上优化了市场准入的制度成本。在基础模型规制路径上，欧盟侧重于事前合规审查，美国则注重行业自律，体现出双方在治理逻辑上的差异化侧重。

亚太经合组织(APEC)在数字经济框架下的治理实践呈现出区域特征。其跨境隐私规则(CBPR)体系致力于构建数据流动信任机制，依托认证制度统筹各成员的数据保护标准。在人工智能领域，APEC 侧重赋能中小企业，通过发布实施指南提供场景化的合规参考。相较于成文法路径，APEC 更多通过各经济体自主采纳的软性规范发挥作用，如新加坡据此制定行业自律守则，日本将其纳入政府采购考量。此种柔性治理模式契合了技术快速迭代的现实诉求，其实际效能有赖于各经济体的协同推进。未来，该机制在算法公平性救济等维度的规则衔接仍有探讨与深化的空间。

金砖国家在人工智能伦理机制建设上正逐步凝聚共识。2023 年出台的相关伦理框架确立了“包容性增长”原则，倡导技术演进应契合新兴市场的特征。各方在数据治理议题上展现出相近的政策偏好，注重统筹数据安全与跨境流动的平衡。巴西据此推进了个人数据保护法规的修订并增设 AI 条款；南非则在公共部门算法应用中引入了社会影响评估程序。此种合作模式充分尊重各国的政策空间，允许基于不同发展阶段设定差异化的合规基线。在特定安全应用等领域，各方仍立足本国实际保留了充分的自主规制权限。[27]依托联合研究中心等机制安排，相关伦理原则向技术标准的实质性转化正持续推进。

表 5 全球 AI 治理多边合作实践对照表

主体名称	协作重点	治理模式	典型成果	机制特点
欧盟-美国 TTC	技术标准互认；风险管理协同；算法透明度框架	双方存分歧(欧盟事前审查；美国自律主导)	非约束性生物识别技术指南	定期对话工作组机制
亚太经合组织(APEC)	跨境隐私规则；中小企业能力建设；AI 伦理实施	柔性规范(成员自主采纳)	CBPR 体系；《AI 伦理原则实施指南》	依赖成员监管机构配合
金砖国家	AI 伦理准则；数据主权平衡；分阶段合规	保留政策空间(分阶段合规)	2023《AI 伦理框架》；联合研究中心	实体机构推动准则转标准

3.立法协调中的技术壁垒与法律冲突化解

在构建跨国人工智能治理框架的过程中，技术规范层面的差异形成了首要的制度调适课题。不同

[27] 王英明, 石超. 进攻性人工智能的道德考量与全球治理[J]. 科技导报, 2025, 43(4): 37-45.

司法管辖区通常基于各自的产业基础与技术路线，确立相互独立的技术标准体系。以欧盟推行的强制性算法透明度要求为例，其具体参数设定与北美地区侧重行业自律的指导性规范存在差异，这使得面向全球市场的产品在合规成本上面临挑战。由于国际互认的测试认证机制尚在探索阶段，同一人工智能系统往往需应对多重标准的重复评估，这在客观上对产品的研发周期与创新效能提出了更高要求。此外，数据接口协议等核心技术兼容性问题，也是影响跨国协作项目实施进展的客观因素。

关于人工智能系统法律地位的确立，不同法系展现出多元的制度观察。大陆法系通常遵循“主体-客体”二元结构，对赋予人工智能独立法律人格保持审慎，侧重于将责任归于开发者或使用者；部分普通法系背景下的讨论则在探索针对具备自主决策能力的系统，确立更具针对性的责任归属机制。这种认知层面的差异在自动驾驶跨境事故等场景中尤为显著，不同法域对相关技术特性的法律定性，可能对裁判结果的一致性产生影响。这种现象不仅折射出法律传统的差异，也反映了不同文化背景下对技术风险感知度的差异化。

面对跨境法律冲突的演进，传统冲突法规则的应用路径正经历深度研讨。有观点主张优化“最密切联系原则”，通过加权分析开发地、数据来源地、损害发生地等要素来确定准据法。然而，人工智能技术典型的分布式特征——如云端训练、边缘计算、全球化交互——使得传统连接点的界定更具复杂性。此外，关于构建数字领域统一冲突规范公约的构想，亦涉及国际法治协作中主权协调等深层议题。在数据隐私等特定领域，相关司法实践通过“市场目的地原则”等方式探索域外管辖的适用，虽各界评价不一，但也为解决现实纠纷提供了制度参考。当前，全球范围内化解此类法律冲突的成熟方案仍有待各方在持续对话中共同构建。

四、全球人工智能治理立法的演进趋势与愿景

(一)趋势研判：立法理念的深刻变革

1.从单域治理迈向跨界融合的立体立法

人工智能技术的深度渗透已使医疗、交通、制造及日常生活等领域高度交织，跨场景应用对传统分行业规制模式提出了挑战。当技术要素在不同行业间跨界流动时，原有的条块分割式立法思维在规则覆盖的全面性与衔接的连贯性方面，仍有进一步优化的空间。

构建跨领域的协作机制、打破部门间的信息壁垒已成为制度演进的重要方向。通过设立专门的协调机构或完善联席会议制度，能够有效汇集技术研发、数据安全及行业应用等多方专业力量，共同评估新兴 AI 应用的多维度影响，从而制定兼顾创新活力与风险防控的综合对策。例如，当医疗影像识别模型转化为自动驾驶的环境感知算法时，其涉及的数据合规与责任归属需由医疗、交通及工业信息化等部门协同审视，以确保规则适用的准确性。

未来的治理目标在于构建覆盖技术源头至应用全链条的立体化管理网络，同步聚焦技术特性演进、数据流动保护及场景应用安全。规则制定者需整合科技、经济、社会与伦理等多重因素，推动各领域规范的深度衔接，形成具有高度适应性的治理体系，在保障技术进步动能的同时，实现技术演进与社会治理的协同共生。

2.从地域性标准迈向国际共识标准的构建

全球人工智能治理目前呈现出规则多元化的态势，不同经济体基于各自的法律传统、产业基础及价值取向，确立了差异化的监管框架。欧盟以严格的风险分级为特征，美国强调创新效率与灵活规制，中国则致力于统筹发展与安全。这种地域性标准的并存反映了治理诉求的多样性，同时也对跨国技术研发与产品流通提出了更高的合规要求，影响了全球范围内的产业协同效能。[28]

ISO、IEEE 等国际标准组织在汇聚全球治理共识方面正发挥着不可或缺的作用。相关机构在术语定义、技术规范、测试方法及伦理准则等基础层面积极寻求一致性：ISO/IEC JTC 1/SC 42 聚焦于系统可信赖性等核心维度，IEEE 则致力于推动伦理对齐设计等标准的制定。这些努力为弥合区域分歧、构建通用的技术语言与互操作规范提供了关键路径。

推动地域性标准向国际共识演进，需在多个层面探索协同路径：首先是建立标准互认机制，通过双边或多边框架实现对符合公认基准的认证结果的相互认可；其次是完善跨境测试认证体系，在测试环境、评估指标及数据共享等方面加强衔接；此外，还需进一步凝聚国际伦理共识，围绕人工监督、公平性与隐私保护等核心维度确立基准规范。其最终愿景在于构建一个包容多元且保障技术要素顺畅流动的全球规则网络。

3.从安全底线迈向可持续发展的价值导向

人工智能立法正经历从单一风险防控向可持续发展价值导向的深度转型。早期治理主要聚焦于规避技术失控、数据泄露及算法歧视等底线风险，而随着应用程度加深，资源效能、社会公平及生态影响等议题日益凸显。当前的立法理念正超越被动防御，更加主动地引导技术服务于人类福祉与生态健康。

在立法实践中，绿色低碳原则正被引入人工智能的全生命周期管理，并转化为具体的能效准入、碳足迹测算及环境影响评估机制。通过要求特定规模的系统提交资源消耗报告，旨在推动高能效算法模型与硬件设施的研发应用。

[28] 孙志伟.从技术竞逐到“智能公域”：人工智能全球治理的范式转向[J].国际展望,2026,18(01):40-61+174-175.

包容普惠已成为人工智能法治建设的重要支柱，旨在确保技术红利的公平分配。这集中体现为对公共服务可及性的法律要求，如支持多语言与无障碍交互；同时，在算法公平性层面，规则制定愈发强调纠正数据偏见，并对特定群体予以必要的保护。此外，支持低成本人工智能工具的开发，有助于小微企业与个体经营者共享技术演进成果。

推动人工智能与联合国可持续发展目标(SDGs)的深度融合是当前的核心愿景。通过建立跨领域协同机制，例如将健康 AI 的伦理审查与公共卫生目标挂钩，能够确保技术创新导向社会繁荣与生态平衡的统一，从而实现技术潜力与社会价值的和谐共生。

(二)路径优化：治理方式的敏捷性与包容性转型

1.引入“监管沙盒”提升制度的创新相容性

监管沙盒作为人工智能治理的探索性工具，旨在通过构建受控的真实环境，在确保风险可控的前提下，为技术创新提供制度试错空间。其核心价值在于通过近距离观察技术的社会表现与风险特征，推动监管规则的适应性调控。

在国际实践中，英国金融行为监管局率先确立了沙盒模式的运行规范，新加坡金融管理局随后将其拓展至人工智能金融应用领域。其运行机制通常涵盖严密的准入、过程与退出流程。准入阶段侧重评估技术创新性、公共利益贡献度及风险防控方案；测试阶段依托实时的监管监测与数据反馈，保持双方的信息对称；退出阶段则依据测试结果决定该技术的后续规制路径，即具备合规基础的项目将获得准入授权，而风险溢出的方案则需进入迭代优化。

监管沙盒的引入，有效平抑了人工智能快速迭代与法律规则稳定性之间的固有张力。该机制通过提供真实的测试场景，在避免过度规制抑制创新动能的同时，依托前置化的风险识别降低了系统性风险。随着治理经验的积累，进一步提升沙盒运行的普惠性、优化跨部门的协同效能，并将其沉淀为更具普遍适用性的监管准则，正成为该制度演进的关键方向。

2.构建“技管结合”的技术驱动型监管体系

人工智能治理正逐步转向技术驱动模式，其核心在于依托技术手段优化规制效能。监管科技的引入为提升规制穿透力提供了实质支撑。区块链技术的不可篡改特征为解决算法透明度问题提供了路径，通过完整记录算法运行日志与决策参数，能够形成可追溯的审计轨迹，在降低黑箱风险的同时，为责任归属的判定提供事实依据。

人工智能技术本身也被应用于监管的智能化转型。通过机器学习对相关主体运行数据的深度分析，监管系统能够自动识别异常模式或潜在风险点，从而实现从被动响应向主动预警的转变。这种基于算法的监测机制具备显著的敏捷性，能够及时识别风险萌芽并缩短响应周期，例如在金融科技领域对异常数据流的实时监测。结合大数据分析，监管过程演变为持续学习与动态调整的智能研判体系。技术赋能的本质在于构建具备自我演进能力的治理架构，在夯实安全基座的同时，维系产业的创新活力，进而形成监管与发展的良性互动。

3.强化以“人机和谐”为核心的伦理立法前置

人工智能领域的法律规制往往面临技术迭代领先于制度供给的挑战，这导致伦理准则在实际应用中存在滞后感。现有的治理模式多侧重于争议发生后的被动响应，难以在技术研发阶段有效预防系统性风险。因此，将人类尊严与公平正义等伦理价值前置于法律框架之中，已成为当前制度转型的重心。

针对算法偏见，可以探索在立法中要求研发主体在设计阶段嵌入偏差检测程序，并将其作为产品获准入市的合规要件。在医疗诊断等涉及生命安全的领域，则应明确保障专业人员的最终决策否决权。

跨学科伦理委员会的组建是实现该治理目标的关键支撑。这类机构应整合法学、伦理学与计算机专业力量，负责确立伦理评估基准，并引导企业在研发初期开展伦理影响评估。具体而言，审查工作应重点关注医疗算法等应用在数据样本上的多元性，以及决策逻辑是否具备医学上的可解释性，从而在源头规避潜在风险。

动态伦理评估机制的建立需依据场景的风险等级实施差异化管理，即对聊天机器人等低风险应用实施简化流程，而对自动驾驶、金融风控等高风险领域落实全生命周期的监测。相关主体应按期提交运行数据以备核查。法律层面需清晰界定技术失范的判定标准，即当算法逻辑偏离人类主流价值或引发负面效应时，应及时启动强制性干预。此外，人类最终控制权等原则应通过强制性法律条款予以确立，确保技术演进始终服务于人类整体福祉。

(三)愿景展望：构建共治共享的法律文明形态

1.推动立法从博弈防范转向共赢协作

人工智能治理正迈向从防范性博弈到共赢协作的转型阶段。此前，基于技术优势保护的考量，各国在监管层面常出现摩擦，典型的如算法出口限制与数据本地化要求。这些措施在客观上推高了全球合规成本，并对解决气候变化预测模型等跨国协作议题构成了限制。鉴于人工智能技术具有极强的跨境流动性，单一的治理框架难以应对算法偏见等复杂课题，这要求政府、企业与科研机构等多方主体达成深度协作。在制定自动驾驶技术标准等具体场景中，多元主体的参与能够更敏锐地识别潜在的风险盲区。[29]

在实践路径上，构建务实的协作载体已成为各界共识。国际人工智能治理基金的设立有助于支持更多经济体参与规则制定，优化治理结构的包容性。同时，技术共享平台能够减少研发资源的重复投入，医疗影像开源模型的应用便是显著提升全球癌症筛查效能的典型案例。该模式的实质在于确立风险共担与收益共享的分配机制，这涉及隐私保护责任的明确界定以及跨国赔偿基金的筹措。这种转型的实现，有赖于主要经济体在量子计算规制等前沿领域建立互信，并超越短期的竞争思维。当前，多边对话机制正逐步消解传统的博弈定式，为全球协同治理铺平道路。

2.塑造具备韧性与弹性的全球法治协作网络

在人工智能全球治理领域，多元的治理路径在满足不同发展诉求的同时，也伴随着规则衔接的复杂性。技术要素的跨国流动与主权管辖的边界，驱动各方探索具备适应性与韧性的全球法治协作架构。此框架应具备包容技术迭代的能力，其制度设计旨在调适不同发展路径间的张力，维护治理体系的整体稳定性。

联合国及其拟设的人工智能咨询机构具有发挥协调中枢的潜力，其功能的实现有赖于各方共识的深度凝聚。构建具备议程设置、信息共享及立场协调功能的常态化平台是重要路径。该平台应当定期梳理各国立法动态，研判规则衔接点，汇聚专家力量开展风险预警，从而为国际规则的形成提供实质支撑。

跨境执法协作的完善直接关乎协作网络的实效。鉴于人工智能应用的全球属性，建立联合调查、跨境取证以及针对高风险领域的专项监管协作机制，是提升治理效能的必然选择。相关制度安排应当统筹执法效率与数字主权，在维护公共秩序的

[29] 支振锋.全球人工智能立法的重要关切与未来趋向[J].比较法研究,2025,(06):17-36.

同时，促进技术标准、认证体系及风险评估方法的国际互认。此举不仅有助于降低跨国合规成本，更能营造可预期的商业环境，增强治理体系的内生韧性。这一多层次协作网络的愿景，在于塑造一种在动态演进中达成共识并维持系统稳定的能力，确保国际社会能够依托集体行动维护秩序与安全。

3.实现技术赋能与人类价值守护的制度动态平衡

人工智能背景下的人机关系有赖于法律框架的精细化设计。确立人类最终控制权是界定技术边界、维护决策自主性的重要准则，其制度初衷在于释放技术潜能的同时，守护人类尊严。

构建系统化的技术社会影响评估制度具有实质意义。此类评估应当超越单一的技术安全或经济视角，将伦理、社会公平及就业等多元维度纳入考察范围。借助专家、学者及社会公众的广泛参与，评估的科学性与社会认同将得以提升。相关评估结论可直接引导技术研发方向、优化市场准入标准并推动法规的动态修订，从而实现对潜在风险的前瞻性调控。

治理的最终愿景在于构建具备自我调适能力的法律生态。这种富有韧性的制度体系应当随技术演进不断完善，并始终坚持以人为本。在防范技术异化的同时，致力于促进个体能力的拓展与机会的平等。关注边缘群体的权益保障，有助于弥合数字鸿沟并共享发展成果。法律制度应当在技术赋能与人类价值守护之间达成一种可持续的动态平衡。

结 语

全球人工智能治理正步入从局部规制向全球协作转型的深水区。这一进程的演进不仅涉及制度层面的对齐，更承载着人类社会对技术文明秩序的深层构建。面向未来，法治的力量应当被引向塑造具备高度韧性与适应性的全球协作网络，使其能够跨越不同发展阶段与文化范式，在差异中寻求实质性的共识。

人工智能立法应当超越单纯的防范性思维，转而锚定价值赋能的愿景，引导技术在增进人类福祉与维护生态可持续性中发挥实质作用。这种转型体现了治理理念的升华，即不仅关注安全底线的守护，更致力于构建开放、公平且包容的数字生态。在这一愿景下，法律不再是单纯的限制手段，而是转化为赋能创新的制度载体，在确保人类最终控制权的前提下，为前沿技术的良性演进预留制度弹性。

这种面向未来的法治蓝图，有赖于各方在伦理前置、技术驱动以及协同共治层面达成更深层次的契合。当全球治理从分散的行业实践升华为系统的法治文明形态，技术的发展红利将以前所未有的广度惠及不同群体，有效弥合数字鸿沟，促进社会公平。实现这一愿景的过程，是人类智慧与法治精神在数字时代的一次深刻共振，其终极目的在于保障技术始终运行在造福人类的既定轨道之上，共同开创人机和谐的法治未来。