交大法学 SJTU Law Review
No. 6 (2025)

人工智能立法的价值取向与模式比较

郑　戈*

摘要　当前关于全球人工智能立法模式的流行分类，如以中国为代表的国家主导型、以美国为典型的市场主导型以及欧盟的权利导向型，本质上是基于意识形态立场的标签化处理，未能准确反映各国立法的核心驱动力与实践逻辑。理解人工智能立法的差异应聚焦于其内在价值取向。中国的立法实践以“有序创新”为核心价值，旨在平衡技术发展与风险控制，通过阶段性规划引导产业竞争力提升。美国的立法则以“国家安全”为优先考量，其规制逻辑主要服务于维护技术霸权与应对地缘政治挑战。欧盟则展现出鲜明的“风险规避”导向，通过前置性、预防性的严格规则最大限度降低技术应用的社会风险。这种基于价值取向的模式划分，更能揭示立法差异的深层动因，为理解全球人工智能治理格局提供更客观的分析框架。

关键词　人工智能立法　价值取向　有序创新　风险规避　安全本位

一、问题的提出

包括人工智能在内的数字科技不仅重塑着经济结构与生活方式，更对传统法律秩序构成深层挑战。在这一背景下，阿努·布拉德福德(Anu Bradford)于《数字帝国》(*Digital Empires*)中提出的三分框架——国家主导的中国模式、市场驱动的美国模式与权利本位的欧盟模式——代表着欧美学界分析全球人工智能立法的主流范式。〔1〕比如，在一本研究中国数字经济领域的法律和监管措施的著作中，张湖月将中国加强数字经济法治化的措施描述为“crackdown”(打压)，而将欧盟严监管塑造为“权利保护”。〔2〕这种价值预设遮蔽了真实的立法意图和背后的经济现实。事实上，欧盟《人工智能法案》(AI Act，AIA)虽高举权利旗帜，实则以《欧盟运作条约》第114条(内部

* 上海交通大学凯原法学院教授、法学博士。

〔1〕 See Anu Bradford, *Digital Empire: The Global Battle to Regulate Technology*, Oxford University Press, 2023, part. I.

〔2〕 See Angela Huyue Zhang, *High Wire: How China Regulates Big Tech and Governs Its Economy*, Oxford University Press, 2024.

市场条款)为法律基础,核心目标在于规避成员国监管碎片化风险,权利保护仅作为市场整合的副产品。同样,美国联邦层面系统性回避国内人工智能立法,却通过《芯片与科学法案》(H.R. 4346 - CHIPS and Science Act)实施对外技术封锁,其"内松外紧"的二元结构揭示"市场主导"仅是表象,"国家安全"才是实质立法逻辑。这种基于政治价值预设的类型化不仅日益脱离各国立法实践的真实状态,更导致分析框架的自我消解——当欧盟以风险规避之名行技术贸易壁垒之实,而中国通过地方容错机制与渐进式立法平衡创新与安全时,三分法已沦为毫无意义的空洞表达。

要建设性地提出新的类型学框架,我们首先需要分析目前的流行分类错在哪里。国家主导型、市场主导型和权利导向型三分法存在以下显著缺陷:其一,无法解释中国立法中"发展优先"与"风险可控"的动态平衡。2017年《新一代人工智能发展规划》〔3〕明确提出"分三步走"战略:2020年前聚焦技术突破与产业竞争力,2025年初步构建伦理规范体系,2030年方形成"完善的法律法规体系"。这种渐进式立法路径与其说是国家威权的刚性控制,不如称为"以制度弹性换取创新空间"的有序创新。其二,美国的所谓"市场主导型立法",实际上体现为对于国内的人工智能技术和产业不立法,通过有意留白来为美国本土产业发展保持广阔的空间。2025年7月通过的《大而美法案》(H.R.1 - One Big Beautiful Bill Act)在草案阶段曾经包含的第43201条将"不监管"作为法定义务施加给美国所有的州,要求"任何州及其下属政治实体均不得实施任何关于人工智能模型、人工智能系统或自动决策系统的法律或法规"〔4〕。而联邦层面的几乎所有相关立法都是以"国家安全"之名来限制美国眼中的竞争对手(主要是中国)的发展。从《外国投资风险审查现代化法案》(FIRRMA)〔5〕对AI企业的外资审查,到《2021创新与竞争法案》〔6〕以"技术主权"之名强化本土供应链,市场机制始终嵌套于国家安全框架之内。其三,AIA虽高举权利旗帜,其本质却是风险规避导向的防御性立法。该法案以四层风险分类为核心,禁止"不可接受风险"技术(如大规模监控系统、社会评分系统和"操纵人的潜意识"的系统),却未构建促进技术正向发展的激励机制,折射出"恐惧驱动立法"的保守倾向。但不发展才是真正的不安全,如果整个技术、产业和服务都无法自主,那么严格的风险防范机制除了起到偶尔对中美科技巨头处以高额罚款的作用之外,根本无法发挥这种立法所鼓吹的保护公民基本权利的作用。也正是在这样的背景下,欧盟如今正在考虑暂缓《人工智能法》的实施。〔7〕

本文的论证将沿三条脉络展开:首先,以立法文本、政策轨迹与司法实践为经验基底,解构布拉德福德范式的认知偏误,揭示其如何遮蔽各国立法中的实用主义妥协与策略性调适。其次,通过实证分析中美欧代表性立法案例(如中国地方自动驾驶立法试验、美国《算法问责法案》博弈、欧盟AIA的监管金字塔设计),论证"有序创新—国家安全—风险规避"框架对立法动机、工具选择与实践效力的解释力。最后,反思三类模式对全球技术治理的启示:当人工智能日益成为大国竞争的战略支点,立法模式的选择实则是国家在技术革命浪潮中的生存策略——或主动塑造秩序,或被动防御风险,或试图以泛安全化的涉外法治来维系霸权。

---

〔3〕《国务院关于印发新一代人工智能发展规划的通知》,国发〔2017〕35号。

〔4〕119th Congress of the United States, H.R.1 - One Big Beautiful Bill Act, Section 43201, https://www.congress.gov/bill/119th-congress/house-bill/1/text.

〔5〕Foreign Investment Risk Review Modernization Act of 2018, https://home.treasury.gov/sites/default/files/2018-08/The-Foreign-Investment-Risk-Review-Modernization-Act-of-2018-FIRRMA_0.pdf.

〔6〕United States Innovation and Competition Act of 2021, https://www.congress.gov/bill/117th-congress/senate-bill/1260.

〔7〕Daphné Leprince-Ringuet and Miriam Partington, *EU Could Pause AI Act Rollout Amid Industry Backlash: "It's Really Toxic"*, Sifted, May 29, 2025.

表1　全球人工智能相关立法的三种模式

模式	核心价值取向	立法目标	主要立法工具与特点	代表实践与案例
中国模式	有序创新	平衡技术发展与风险控制，提升产业竞争力	渐进式立法、地方试验性规则（如容错机制）、阶段性规划；强调公私合作与制度弹性	《新一代人工智能发展规划》（“三步走”战略）、地方条例（如深圳《人工智能产业促进条例》）
美国模式	国家安全	维护技术霸权，应对地缘政治挑战	国内宽松监管（依赖行业自律与软法）、对外强硬制裁（出口管制与投资审查）；安全例外机制	《芯片与科学法案》（技术封锁）、《外国投资风险审查现代化法案》（外资审查）
欧盟模式	风险规避	最小化社会风险，保障基本权利	前置性预防规则（风险分类）、强制性合规（如影响评估）、伦理化立法；强调监管冗余	GDPR（数据保护影响评估）、AIA（四层风险分级与禁止清单）

二、中国模式：有序创新

中国人工智能治理体系呈现出鲜明的“有序创新”特征，这种模式既非单纯的国家主导，也不是对市场或个人权利的片面强调，而是通过多层次、多领域的制度设计平衡技术创新与风险防控。长期关注中国人工智能相关政策和法律的卡内基国际和平研究院研究员马特·希恩指出：“北京正在引领全球人工智能规制之路，在算法、聊天机器人等的治理方面都提出了开创性的新战略。”有意思的是，希恩注意到中国并不想在立法上争当第一，而是按照自己的节奏和步伐解决技术和产业发展中出现的现实问题，预防真正存在的风险。尽管尚未出台统一的“人工智能法”，但由部门规章、地方性法规与国家级法律共同构成的立体框架已实质形成完整的规范体系。这一体系可依规制目标分为五类：

其一，分类监管AI技术及其应用的部门规章，主要由国家网信办牵头制定，针对不同技术形态细化规则。例如，《网络信息内容生态治理规定》（2019）确立算法推荐服务提供者的主体责任，要求建立人工干预机制；《互联网信息服务算法推荐管理规定》（2021）首次以“算法分类”为基础，禁止利用算法实施价格歧视或诱导沉迷；《互联网信息服务深度合成管理规定》（2022）将人脸生成、语音合成等技术纳入备案管理，并强制标识合成内容；《生成式人工智能服务管理办法》（2023）进一步要求训练数据合法、生成内容符合核心价值观。此类规章以技术特性为监管切口，通过动态更新回应技术迭代，体现“敏捷治理”思路。

其二，地方性产业促进法规，聚焦激发创新活力与经济赋能。上海《促进人工智能产业发展条例》（2022）首创“容错机制”，允许企业在合规前提下开展高风险技术试验；深圳《人工智能产业促进条例》（2023）则设立专项基金支持中小企业算法研发，并建立伦理委员会提供合规指导。此类立法弱化事前审批，强调基础设施共建（如上海布局智能算力集群）、数据开放共享（如深圳公共数据定向授权企业使用），本质是以制度供给降低创新成本。

其三，保障权利与安全的国家法律，构成我国人工智能法律体系的基石。《网络安全法》（2016）确立关键信息基础设施运营者安全义务；《数据安全法》（2021）首创数据分类分级制度，约束算法训练数据获取；《个人信息保护法》（2021）赋予个体算法拒绝权与解释请求权，直接制衡自动化决策。此类法律虽未冠以“人工智能”之名，却通过规范算力载体（网络）、算法燃料（数据）及算法作用对象（个人信息），实现对技术应用的闭环约束。

其四，人工智能伦理规范体系，我国人工智能伦理规范体系的构建并非停留于抽象原则宣示，

而是通过制度设计将伦理要求转化为可操作的合规义务，实现伦理原则与法律规则的功能耦合。这一体系的核心特征在于将“科技向善”“以人为本”等价值理念嵌入技术开发生命周期，形成预防性伦理约束机制。《生成式人工智能服务管理办法》(2023)第4条明确要求提供者和使用者“尊重社会公德和伦理道德”，并具体化为训练数据质量保障、内容过滤机制及防止歧视性输出的技术义务，体现了伦理要求向技术标准的转化路径。〔8〕地方立法进一步强化了伦理审查的操作性，如深圳《人工智能产业促进条例》第21条规定企业设立伦理委员会“对数据处理、算法设计及应用部署进行伦理评估”，将伦理治理主体内化于企业组织架构。值得注意的是，中国模式与欧盟的“伦理法律化”路径存在本质差异：AIA将HLEG《可信人工智能伦理指南》中的七项原则直接转化为高风险系统的强制性合规清单，试图通过刚性规则穷尽伦理风险；而中国则更强调“伦理治理的场景适配性”，通过《互联网信息服务算法推荐管理规定》第11条要求算法推荐服务提供者“建立完善人工干预和用户自主选择机制”，在电商、社交、内容推送等具体场景中落实“人在回路”(Human-in-the-loop)原则，确保算法决策的最终控制权保留于人类主体。〔9〕这种基于场景的伦理实施机制，既避免了欧盟式合规负担对技术迭代的抑制，亦超越了美国依赖企业自律的伦理空洞化倾向，构成“有序创新”在价值维度的重要体现。

其五，是以人为本的人工智能行业标准体系，行业标准作为技术规则与法律规范的桥梁，在中国人工智能治理体系中承担着将“以人为本”理念转化为技术参数的枢纽功能。我国在《国家人工智能产业综合标准化体系建设指南(2024版)》中表示将持续提升标准与产业科技创新的联动水平，拟新制定国家标准和行业标准50项以上、参与制定国际标准20项以上，引领人工智能产业高质量成长与促进人工智能产业全球化发展。〔10〕地方实践则凸显标准对产业创新的引导作用，《上海市促进人工智能产业发展条例》第66条首创“人工智能伦理专家委员会”制度，其职能就包括“组织制定人工智能领域伦理规范指南”。这种“试点—推广”的标准形成机制，显著区别于美国国家标准与技术研究院(NIST)《人工智能风险管理框架》〔11〕的自愿性指引模式，也不同于欧盟网络安全局(ENISA)标准对AIA合规条款的机械转译。〔12〕更具特色的是中国标准体系的“生态主义”转向：《深圳经济特区人工智能产业促进条例》第28条要求“市产业主管部门应当完善数据中心运行评估评价体系，探索提升数据中心能效标准，搭建支撑人工智能发展的绿色数据中心”，在实施过程中将碳足迹追踪、能源效率等指标纳入评标体系。这表明中国人工智能标准体系已超越单纯的技术安全范畴，向促进技术以增益人类福祉、平衡创新效能与社会价值的系统性规则演进，为全球人工智能治理提供了“标准—法律—产业政策”三维联动的中国方案。

从技术要素视角来看，中国的相关立法也已经全面覆盖人工智能的三大支柱：一是算力层面，《关键信息基础设施安全保护条例》(2021)强制数据中心安全评估，“东数西算”工程通过政策倾斜优化算力布局；二是算法层面，前述部门规章就是针对不同类型的人工智能算法，形成从研发备案、部署透明到结果追溯的全周期规制；三是数据层面，已经形成从《网络安全法》《个人信息保护法》和《数据安全法》到《数据出境安全评估办法》(2022)及地方数据条例(如《浙江省公共数据条

〔8〕 参见郑戈：《人工智能伦理的机制设计》，载《中国法律评论》2025年第1期。

〔9〕 参见郑戈：《人工智能与法律的未来再思考》，载《数字法治》2023年第3期。

〔10〕 参见高秦伟：《人工智能标准规制的监督机制》，载《郑州大学学报(哲学社会科学版)》2025年第3期。

〔11〕 The NIST AI Risk Management Framework (AI RMF), https://www.nist.gov/itl/ai-risk-management-framework. NIST是美国国家标准与技术研究院(National Institute of Standards and Technology)的简称，该机构直属于美国商务部。

〔12〕 参见于连超：《论强制性国家标准的技术法规化变革》，载《电子知识产权》2023年第6期。

例》)构建数据利用与流通的合规路径。

在审视全球人工智能治理图景时,日本近期推出的《关于推进人工智能相关技术的研究开发及活用的法律案》(以下简称"日本AI推进法")〔13〕为我们理解"有序创新"模式提供了另一个重要例证。该法案虽尚在立法进程中,但其核心架构与立法逻辑清晰展现了与中国路径相似的、旨在平衡创新发展与风险规制的"有序创新"特质。该法案开宗明义,指出其目的在于通过综合且有计划地推进AI技术研发与应用,提升国民生活水平并促进国民经济健全发展。为实现此目标,法案构建了系统的治理框架:首先,确立了明确的基本理念,强调AI作为经济社会发展基础技术的重要性,要求国家在确保研究开发能力、提升产业国际竞争力的同时,必须坚持以人为本的应用方向,防范犯罪利用、隐私泄露、侵权等风险,并确保过程的透明性。这与中国在鼓励AI产业发展的同时,通过《网络安全法》《数据安全法》《个人信息保护法》等筑牢安全基石的思路高度一致,均体现了发展优先但不忘安全底线的有序内核。其次,法案设立了强有力的统筹协调机制——"人工智能战略本部",由内阁总理大臣亲自领导,全体国务大臣参与,负责制定并推进实施国家层面的"人工智能基本计划"。这种顶层设计与高位协调机制,类似于中国在AI治理中由中央网信办牵头协调多部委的模式,旨在克服部门分割,实现政策的综合性与计划性,确保创新方向与国家战略目标一致。再者,法案详细规定了国家、地方公共团体、研究开发机构及企业等多元主体的责任与协作义务,要求国家制定综合计划并完善基础设施,地方因地制宜施策,研究机构积极研发与培养人才,企业则需在业务中积极应用并配合国家政策。这种分层治理与协同共治的模式,也非常类似于中国在中央统一规制框架下,鼓励地方(如上海、深圳)探索产业促进立法,并调动市场主体积极性的实践。尤为值得注意的是,日本法案虽以"推进"为名,但并未忽视规制,其"适正性确保"条款要求国家采取措施确保AI研发与应用的正当实施,包括遵循国际规范制定指南等,这与中国针对生成式AI、推荐算法、深度合成等特定技术形态出台专门部门规章进行分类监管的思路异曲同工,均是在"促进"的主基调下嵌入必要的规制模块。因此,日本AI推进法的核心特征——国家主导的顶层规划、发展导向与风险防范并重、多元主体协同共治——有力印证了"有序创新"作为一种超越意识形态的、具有实际效能的治理模式,在东亚法治语境下的广泛适用性与生命力。

表2　中国的人工智能法体系

类　型	主要内容	代表性文件/例子	核心特征
按技术特性进行分类监管	分类监管AI技术及应用,针对不同技术形态细化规则,强调主体责任和动态调整。	以部门规章为主,比如《网络信息内容生态治理规定》(2019)、《互联网信息服务算法推荐管理规定》(2021)、《互联网信息服务深度合成管理规定》(2022)、《生成式人工智能服务管理办法》(2023)	敏捷治理:以技术特性为监管切口,通过动态更新回应技术迭代,禁止算法滥用(如价格歧视、诱导沉迷)。
地方性产业促进法规	激发创新活力与经济赋能,提供制度支持以降低创新成本。	以地方性法规为主,比如上海《促进人工智能产业发展条例》(2022):首创"容错机制";深圳《人工智能产业促进条例》(2023):设立专项基金和伦理委员会	弱化事前审批:强调基础设施共建(如智能算力集群)、数据开放共享(如公共数据定向授权),以制度供给促进创新。

〔13〕《人工知能関連技術の研究開発及び活用の推進に関する法律案》,全文参见日本众议院官网,https://www.shugiin.go.jp/internet/itdb_gian.nsf/html/gian/honbun/houan/g21709029.htm。

续　表

类　型	主 要 内 容	代表性文件/例子	核　心　特　征
维护国家和个人权利	保障权利与安全，规范算力载体、算法燃料（数据）及算法作用对象（个人信息）。	以法律为主，比如《网络安全法》（2016）、《数据安全法》（2021）、《个人信息保护法》（2021）	闭环约束：虽未冠名“人工智能”，但覆盖技术全要素（算力、算法、数据），赋予个体算法拒绝权与解释请求权。
伦理规范体系	将伦理要求转化为合规义务，强调“科技向善”“以人为本”，嵌入技术开发生命周期。	《生成式人工智能服务管理办法》（2023）第 4 条（要求尊重社会公德）、深圳条例第 21 条（设立企业伦理委员会）	场景适配：预防性伦理约束，注重“人在回路”原则（如人工干预机制），区别于欧盟的刚性合规清单。
行业标准体系	行业标准作为技术规则与法律规范的桥梁，促进产业创新与全球协作。	《国家人工智能产业综合标准化体系建设指南》、上海条例第 66 条（伦理专家委员会制定指南）	生态主义转向：涵盖能效标准（如碳足迹追踪），形成“标准—法律—产业政策”三维联动模式。
技术要素视角	覆盖人工智能三大支柱（算力、算法、数据），实现全周期规制。	算力：《关键信息基础设施安全保护条例》（2021）、“东数西算”工程；算法：部门规章全周期管理；数据：《数据出境安全评估办法》（2022）、《浙江省公共数据条例》	全面覆盖：通过政策倾斜优化算力布局，构建数据利用与流通的合规路径，确保技术可控性。

三、欧盟模式：风险规避

欧盟人工智能立法的核心特征并非其宣称的“权利本位”，而是植根于其法律传统与制度约束的风险规避模式。这一模式通过成熟的监管范式移植、精细的风险等级划分以及技术化的执行机制，构建了一套旨在系统性预防人工智能潜在危害的防御性框架，其内在逻辑优先服务于欧盟内部市场统一与经济协调，权利保护则构成其表面目标。

图一　欧盟AI Act — 以风险治理为导向　©郑戈

合理性论证　欧盟人工智能法案的核心理念在于通过“以人为本、风险预防”原则构建可信赖的人工智能生态系统，其主要目的是在确保基本权利、安全性和伦理准则的前提下，以风险分级监管为关键手段（包括禁止不可接受风险系统、严格规范高风险系统、强化透明度和问责机制），推动人工智能技术的负责任创新与应用。

1. 风险规避模式的萌生与成熟：GDPR

欧盟人工智能立法的核心逻辑植根于其独特的风险规避路径，这远非简单的“权利导向”标签所能概括。该模式的形成始于《通用数据保护条例》(GDPR)的立法实践(2012—2016)，其本质是将抽象的权利保障诉求转化为可操作的风险管理技术。拉斐尔·盖勒特(Raphaël Gellert)的研究揭示了这一转化机制：〔14〕欧盟立法者重构了“比例原则”，将传统权利平衡(如干预措施的“必要性”审查)转化为技术化的风险评估流程(如数据保护影响评估DPIA)。〔15〕GDPR第35条引入的DPIA制度，〔16〕要求企业根据数据处理活动的风险等级配置差异化合规措施，表面上是产业界倡导的“灵活合规”方案，实际上则是通过“鉴于条款”第76条(Recital 76)将其定位为“实现比例原则的技术手段”，使风险评估既成为企业的管理义务(GDPR第24条)，又构成监管机构的执法标尺。这种设计虽以权利保护为名，却埋下了概念置换的隐患：GDPR将“数据保护风险”(实体权利侵害可能)悄然替换为“合规风险”(监管制裁概率)，导致企业资源集中于可量化的罚款规避(如数据泄露)，而忽视难以货币化的系统性风险(如算法歧视)。欧盟后续立法尝试修正此偏差，但其风险规避的底层逻辑已深刻定型。

AIA将风险规避模式推向制度化顶峰，其核心特征在于对成熟产品安全监管范式的系统性移植。受限于欧盟权能分配体系[尤以《欧盟运作条约》(TFEU)第114条内部市场条款为法律基础]，AIA无法以基本权利保护作为独立立法依据，被迫将AI系统首要界定为可能危害健康、安全及基本权利的“危险产品”[Article 1(1)AIA]，权利保护仅作为市场整合的副产品。这直接塑造了AIA的监管架构：其一，沿袭“风险分级—技术标准—合格评定—市场监督”的产品安全链条，高风险AI系统(列于Annex Ⅲ)〔17〕需满足强制性要求(如风险管理、数据质量、技术文档、透明度)，其具体化依赖欧洲标准化组织(CEN/CENELEC/ETSI)主导的技术标准(Article 40 AIA)，企业符合标准即可获得“合规推定”。这延续了GDPR以来的将技术风险转化为可控的“合规风险”的路径；〔18〕其二，执行机制依赖“自我控制”(企业自评或公告机构评估，无须行政事前许可)和“事后监督”(成员国市场机构依《市场监管条例》进行)，形成形式上高效但实质松散的“元规制”模式；其三，严格限定高风险范围(如允许企业以“不构成显著风险”为由自我豁免义务)，并为占市场主体的非高风险系统预留宽松空间(仅需有限透明度义务或自愿守则)，法案明确禁止成员国对此类系统施加额外限制(AIA作为“监管盾牌”)。这种设计虽宣称平衡创新与权利，实则暴露其优先目标在于规避成员国监管碎片化以保障内部市场运行，权利保障深度受制于欧盟内部市场一体化和贸易自由化需求。

2. 欧盟人工智能法的立法逻辑：产品安全监管范式向人工智能产业的移植

AIA的立法基本思路清晰展示了其风险规避的本质。法案首要法律依据选择TFEU第

---

〔14〕 Raphaël Gellert, *The Risk-Based Approach to Data Protection*, Oxford University Press, 2020, p.166－170.

〔15〕 DPIA是适用于数据处理相关的风险管理的重要规制工具，关于这一工具的起源和发展，请参见Roger Clarke, *Privacy Impact Assessment: Its Origins and Development*, 25 Computer Law & Security Review 123 (2009).

〔16〕 参见《通用数据保护条例》的评注书：Christopher Kuner, Lee A. Bygrave & Christopher Docksey eds., *The EU General Data Protection Regulation (GDPR): A Commentary*, Oxford University Press, 2020, p. 665－679.

〔17〕 EU, Artificial Intelligence Act, Annex III: High-Risk AI Systems Referred to in Article 6(2), https://artificialintelligenceact.eu/annex/3/.

〔18〕 See Raphaël Gellert, *Understanding the Notion of Risk in the General Data Protection Regulation*, 34 Computer Law and Security Review 279 (2018).

114 条(内部市场条款),而非更直接关涉权利的 TFEU 第 16 条(数据保护条款)。这一选择并非偶然,其核心目的在于规避成员国监管碎片化、统一内部市场规则、保障技术自由流动,权利保护仅作为次要目标嵌入其中[Recital 5,Article 1(1) AIA]。法案将人工智能系统首要定性为可能危害健康、安全、环境、基本权利、民主与法治的"危险产品",而非以公民权利保障为绝对核心构建。正如一位学者所言:"《人工智能法案》从产品安全立法的视角对人工智能系统进行规制:无论公共部门抑或私营部门使用的人工智能系统,均被界定为对健康、安全、环境、基本权利、民主及法治具有潜在危害性的产品(依据该法案第 1 条第 1 款)。基于 TFEU 第 114 条之规定,必须在欧盟层面协调相关监管规则,旨在确保内部市场的正常运转,同时避免成员国制定限制性国内规则阻碍人工智能系统在欧盟范围内的自由流通。"〔19〕

为有效实现风险规避,AIA 系统性地移植了欧盟成熟的产品安全监管架构,构建了"风险分级—技术标准—合格评定—市场监督"的四步监管链条:(1) 风险分级识别:依据潜在危害程度对 AI 系统进行分类。(2) 技术标准具象化:主要依赖欧洲标准化组织(CEN,CENELEC,ETSI)制定技术标准(Article 40 AIA),辅以欧盟委员会制定的通用技术规范(Article 41 AIA),将抽象要求(如数据质量、鲁棒性)转化为可验证指标。符合标准或规范可获得"合规推定"。(3) 合格评定:高风险系统提供者需通过"自我控制"(企业自评或公告机构评估,无须行政事前许可,Article 43,46 AIA)证明符合性。(4) 市场监督:成员国监管机构依据《市场监管条例》(EU 2019/1020)进行事后监督(Articles 70,74 AIA),并与特定行业安全法规(如机械、医疗器械指令)相协调(Articles 102 - 109 AIA)。这一基于既有产品安全工具箱的模式,旨在系统性规避 AI 技术失控带来的广泛社会风险,其出发点在于市场秩序与安全,而非基本权利关切。

3. 核心机制:严格风险分级与靶向规避

风险规避的核心理念在 AIA 中直观体现为对 AI 系统的严格风险等级划分及其对应的差异化监管强度:首先是从源头禁止"不可接受风险":法案明确禁止八类被视为具有不可接受风险的 AI 实践(Article 5 AIA),包括利用潜意识技术、利用特定群体脆弱性、社会评分、执法中的非事实基础预测性警务、大规模无方向性网络抓取生物识别数据创建数据库、特定敏感场景(教育、职场、执法)的情绪识别及生物特征分类系统,以及公共场所的实时远程生物识别(但有严格例外)。这些禁令直指最可能侵犯基本权利、助长监控或歧视的典型高风险应用,旨在从源头规避最恶劣后果。其次是高强度约束"高风险"系统:高风险系统(主要列于 Annex Ⅲ,涉及关键基础设施、教育、就业、公共福利分配、执法司法、移民管控等八大领域)被施加强制性合规要求(Articles 8 - 15 AIA),包括建立风险管理系统、确保数据质量与治理、保存技术文档与记录日志(Logging)、保障透明度与人机协同、维持准确性及网络安全等,并需注册于欧盟数据库(Article 71 AIA)。再次是显著限制高风险范围与预留创新空间:高风险范围的认定极具限制性。Annex Ⅲ所列独立 AI 系统类别有限且集中于公共部门关键决策场景;系统提供者可自行评估其系统是否"不构成显著风险"(如不实质性影响决策结果,满足特定条件)以豁免高风险标签[Article 6(3) AIA],仅需备案[Article 6(4) AIA]。占市场绝大多数的非禁止、非高风险 AI 系统(有限风险、最小风险),仅需遵守有限的透明度义务(如告知用户正与 AI 互动、内容为 AI 生成,Article 50 AIA)或自愿性行为准则(Article 95 AIA)。法案明确禁止成员国对此类系统施加额外限制(AIA 作为"监管盾牌"),实质上为低风险 AI 创新预留了广阔空间。此设计表明,法案的首要目标并非全面限制 AI 发展,而是

〔19〕 Oriol Mir, *The AI Act from the Perspective of Administrative Law: Much Ado About Nothing?*, 16 European Journal of Risk Regulation 63,65 (2025).

精准规避特定高风险领域。

AIA对公共部门应用的规制也体现了风险规避逻辑下的有限进展与显著局限。法案并未设立专门章节规范公共行政自动化决策(ADM),也未直接回应行政法学界长期关注的诸多核心议题,如自动化决策的合法性基础、算法的法律性质、自由裁量中的AI应用、机器学习系统决策理由说明的困境、司法审查前的人工复审机制、利害关系人获取算法及源代码的权利,以及专门的损害赔偿责任制度。其对受AI影响的自然人和法人的权利保障也未构建完整的权利清单。法案对公共部门的主要规制仍嵌套于其风险分级框架内。

综上,AIA通过移植产品安全监管范式、实施严格风险分级、聚焦高风险领域禁令与合规要求并为低风险创新松绑,清晰地勾勒出欧盟"风险规避"模式的轮廓。其对公共部门AI应用的规制虽引入若干重要义务(如影响评估、注册、解释权),但在范围、深度与时效性上均受制于其核心立法逻辑与妥协,本质上提供了一套以规避系统性风险、保障内部市场运行为优先的"最小公分母"式框架,而非以全面保障公民权利或精细规制行政自动化决策为核心诉求的"权利导向"型立法。成员国后续的补充性立法将成为检验其实际风险控制效果的关键变量。[20]

4. 伦理先行

高级别专家组(High-Level Expert Group, HLEG)作为AIA起草中的关键力量,通过其《可信赖人工智能伦理指南》[21]及《政策建议》[22]构建了"风险伦理化"向"法律化"转化的核心路径。其作用体现为三重递进机制:首先,HLEG将抽象风险转化为伦理框架。基于欧盟《基本权利宪章》,提出"可信赖AI"的三元结构——合法性(遵守现行法)、伦理性(动态价值遵循)、稳健性(技术与社会双重保障),并且提炼出四大伦理原则(人类自主、无害、公平、可解释性)与七项关键要求(人类主体性与监督;技术稳健性与安全;隐私与数据治理;透明度;多样性、非歧视与公平性;社会与环境福祉;可问责性)。这一框架突破传统技术风险范畴,将"风险"定义为"系统性社会危害",涵盖个体权利侵害(如歧视)、制度性侵蚀(如民主进程扭曲)及环境福祉威胁,实现风险的"伦理化"重构。其次,通过操作化工具推动伦理向规范过渡。HLEG设计了《可信赖AI评估清单》(ALTAI),将七项要求转化为具体行动问题,要求开发者全生命周期评估伦理可接受性,强调"无伦理妥协方案时禁止部署"。其"风险分级"理念主张以"危害强度与概率"划分监管层级,并引入预防原则应对"不可接受风险"(如生物识别监控),为法律分类提供雏形。最后,AIA选择性吸纳并重构HLEG伦理框架:"可信赖AI"被写入立法目的(第1条),但剥离其三元性,仅保留"以人类为中心"标签;七项要求被援引为"规范性支柱"(Recital 27),却弱化为非强制原则。采纳分级逻辑但窄化风险定义,将HLEG的"社会系统风险"压缩为个体权利威胁,形成"禁止-高风险-可接受风险-低风险"四级体系。例如,人类监督(第14条)、数据治理(第10条)等条款直接源于HLEG要求,但仅限"高风险"场景,且以"封闭式清单"(附件三)取代原则性判断,导致系统性风险(如环境影响)被排除。此外,HLEG主张的"基本权利影响评估"(覆盖所有公私营重大影响系统)被限缩为少数公共部署者义务(第27条);其倡导的"利益相关者多方参与"和"外部审计"仅通过软性行

〔20〕 Ibid.

〔21〕 High-Level Expert Group on AI, *Ethics Guidelines for Trustworthy AI*, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

〔22〕 High-Level Expert Group on AI, *Policy and Investment Recommendations for Trustworthy Artificial Intelligence*, https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence.

为守则(第 95 条)及有限咨询协商(第 67 条)部分实现。[23]

HLEG 的伦理框架本质上是一种风险预分配机制。它将抽象人权话语转化为具体技术指标，如将“公平性”量化为算法偏差检测标准，将“问责”具象为日志记录与审计接口规范。这种转化使伦理原则脱离哲学争论场域，嵌入产品设计生命周期。欧盟委员会在 AIA 提案中明确承认：“HLEG 的工作为高风险 AI 系统的分类提供了科学依据。”[24]法案第 6 条将八类应用场景列为高风险领域，其中招聘筛选、信用评估、执法预测等场景的划分直接源于 HLEG 对“社会关键领域”的界定。更关键的是，法案第 15 条要求高风险系统设立风险管理系统、数据治理机制和自动化日志——这正是 HLEG 技术稳健性原则的法规化表述。

5. 欧盟宪制结构对风险规避模式的强化

AIA 采用产品安全立法模式的核心原因在于欧盟权能分配体系的宪法性约束。这种约束具体体现为欧盟法律体系中的“权能划分原则”(principle of conferral)，即欧盟仅能在成员国授予的有限权能范围内行事。该原则在《欧洲联盟条约》(TEU)第 5 条第 2 款明确规定：“欧盟仅在成员国通过条约赋予的权能范围内行动。”具体而言，欧盟缺乏独立的基本权利立法权能，尽管《欧盟基本权利宪章》具有法律约束力，但其第 51 条第 2 款明确规定：“宪章不扩展欧盟权能范围。”这意味着基本权利保护本身不能作为独立的立法依据。欧盟若以保护基本权利为名立法，需证明该立法与既有权能(如内部市场)存在直接关联，否则可能被欧洲法院判定越权。[25] AIA 的主要法律基础是 TFEU 第 114 条(内部市场条款)，该条款授权欧盟通过“协调成员国法律”以消除市场壁垒。但该条款的适用需满足两个条件：(1) 立法需以“建立或维持内部市场功能”为主要目标；(2) 基本权利保护只能是“附带效果”而非“独立目标”。[26] 因此，AIA 必须被包装为“消除 AI 规则国别差异对内部市场的阻碍”，而非单纯的权利保护工具。

欧盟的宪制结构还直接导致了三大机构的价值导向差异所造成的风险厌恶型妥协。AIA 的立法进程表现出三大机构间权能分配与价值导向的结构性张力。欧盟委员会作为提案主体，严格遵循 TFEU 第 114 条内部市场条款的法律基础，将法案定位为消除成员国监管碎片化的市场工具。这一权能选择决定了法案的核心逻辑：风险分级监管本质上服务于统一产品安全标准的经济目标，基本权利保护仅作为衍生性要求(Recital 5)。委员会在附件Ⅲ中列明八类高风险场景时，刻意排除执法、边境管理等敏感领域，反映出对成员国主权的妥协倾向。此种设计暴露了权能约束下的制度悖论：当立法者试图将人脸识别系统归类为“高风险”时，必须证明该分类与内部市场功能的直接关联，而非基于其对隐私权的实质侵害。欧盟理事会作为成员国利益代表，进一步强化了风险规避路径的防御性特征。其推动的“纯辅助性”例外条款[Article 6(3)]允许企业通过自我评估规避高风险义务，实质弱化了监管确定性。在关键领域豁免谈判中，理事会成功将实时远程生物识别技术排除出禁止清单(仅保留公共场所的有限限制)，并为移民管理系统设置长达六年的过渡期(至 2030 年)。此类妥协导致法案形成双重标准：私营领域的招聘算法需满足全面透明度

---

[23] Nathalie A. Smuha, *The Work of the High-Level Expert Group on AI as the Precursor of the AI Act*, in Ceyhun Necati Pehlivan, Nikolaus Forgó & Peggy Valcke eds., AI Governance and Liability in Europe: A Primer, Wolters Kluwer, 2025, chapter 3.

[24] Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final.

[25] Case C-376/98, Germany *v* Parliament and Council, [2000] ECR I-8419.

[26] Case C-491/01, British American Tobacco (Investments) and Imperial Tobacco, 2002.

要求，而执法机关的同类型算法仅需最低限度披露。更深远的影响在于，理事会坚持将国家安全完全排除出法案适用范围[Article 2(3)]，使军事、情报系统彻底脱离监管框架，印证了风险规避逻辑在主权议题面前的局限性。

欧洲议会虽以基本权利捍卫者自居，但其人权诉求最终被嵌套于产品安全框架之内。议会曾主张采用TFEU第16条（数据保护条款）作为法律基础，试图将算法透明度、非歧视等原则确立为独立权利，但该动议因违背权能分配原则遭否决。其推动的“基本权利影响评估”（FRIA）虽被纳入法案（Article 27）〔27〕，却在实施机制上呈现形式化特征：评估标准由欧洲标准化组织（CEN/CENELEC）制定，而该机构受产业资本主导且缺乏人权法专业能力。议会倡导的个体申诉权（Article 78）同样受制于执行架构——申诉需经成员国市场监管机构过滤，公民无法直接诉诸司法救济。这种权利保障的间接性，暴露了欧盟模式的根本矛盾：当伦理原则必须转化为可量化的技术参数（如算法偏差阈值），人权判断便退化为合规清单中的复选框。

三大机构的博弈最终催生出一部“权利外衣下的产品安全法”。委员会的市场整合优先主义、理事会的主权防御倾向与议会的伦理理想主义，共同将法案压缩为权能公约数下的折中产物。其结果既印证了风险规避模式的技术治理本质，亦揭示了欧盟数字立法的深层困境：在缺乏独立人权立法权能的宪法约束下，任何试图通过产品安全框架承载权利保障的尝试，都可能陷入制度性自我消解。〔28〕

四、美国模式：安全至上

美国在人工智能领域的立法模式呈现为典型的“安全至上”范式，其本质是以消极立法姿态维护技术霸权，通过安全例外机制实现战略遏制。布拉德福德将美国归类为“市场主导型”模式虽反映了美国国内产业发展的客观状态，却未能揭示其立法逻辑的核心矛盾：市场繁荣恰是联邦立法系统性缺位的产物，而非积极政策选择的结果。联邦层面迄今未通过综合性人工智能监管法案，这种“监管真空”绝非偶然，而是刻意为之的制度设计。其内在机理在于通过豁免平台责任（如《通信规范法》第230条）、弱化反垄断执法及维持数据自由流通，最大限度降低本土企业合规成本，从而巩固美国在全球人工智能产业链的支配地位。这种“不作为立法”构成市场优势的制度基础，但本质上服务于技术霸权的战略目标。

1. 美国国内人工智能技术和产业发展：不立法

美国在人工智能领域的立法模式常被归类为“市场主导型”，但这种标签化描述未能揭示其内在逻辑。联邦层面系统性回避国内数据、算法和算力立法的事实，构成美国模式最显著的特征。作为全球人工智能技术最发达的国家，美国迄今未制定全国性的个人数据保护法或综合性人工智能监管法律，这种“监管真空”绝非偶然的制度缺位，而是刻意为之的战略选择。其核心逻辑在于通过最小化国内监管负担，维持本土企业的全球竞争优势，从而服务于技术霸权的国家安全目标。

美国联邦政府始终避免对人工智能实施具有普遍约束力的立法约束。白宫发布的《人工智能

〔27〕 Article 27: Fundamental Rights Impact Assessment for High-Risk AI Systems, EU Artificial Intelligence Act, https://artificialintelligenceact.eu/article/27/.

〔28〕 Francesca Palmiotto, *The AI Act Roller Coaster: The Evolution of Fundamental Rights Protection in the Legislative Process and the Future of the Regulation*, 16(2) European Journal of Risk Regulation 770-793 (2025), doi: 10.1017/err.2024.97.

权利法案蓝图》(2022)声明其“并非美国官方政策,不具备约束力”,仅作为“指导人工智能系统设计、使用及部署”的自愿性建议。[29] NIST 制定的《人工智能风险管理框架》(2023)[30]同样属于技术指南性质,企业可自主决定是否采纳。这种“软治理”模式依赖行业自律与分散化的州级立法,例如纽约市《自动就业决策工具法》(Local Law 144 of 2021)仅规范招聘算法偏见,康涅狄格州《数据隐私法》赋予用户退出自动化决策的权利,但均未形成全国性统一规则。联邦层面的立法克制直接导致监管碎片化:50 个州可能衍生 50 套不同的 AI 监管标准,企业需承担高昂的合规成本,但大型科技公司凭借资源优势可游说各州维持宽松环境。[31]

这种消极立法姿态的本质是以制度弹性换取创新空间。美国国会多次搁置《算法问责法案》《面部识别技术禁令法案》等提案,其根本考量在于避免刚性规则抑制技术迭代速度。正如美国商务部在《人工智能风险管理框架》说明中所强调:“过早的联邦监管可能阻碍新兴技术探索。”[32]硅谷资本对此策略的推动力不容忽视:风险投资机构通过政治献金与游说集团(如 TechNet)施压立法机关,确保监管政策倾向“创新友好型”设计。微软、谷歌等企业在拜登政府《关于安全、可靠和可信赖地开发和使用人工智能的行政令》(EO 14110)[33]起草过程中,成功将强制安全测试改为自愿承诺机制。这种“资本—政策”闭环使市场力量成为事实上的监管主体,将社会成本转嫁给公众:算法歧视诉讼依据 1964 年《民权法案》提起,消费者隐私依赖 1970 年《公平信用报告法》保护,法律工具严重滞后于技术现实。

美国联邦层面在人工智能国内治理领域呈现出显著的立法克制特征,这种“市场主导”表象实则是国家安全逻辑下的制度设计。H.R.1 法案(即《大而美法案》)作为美国近年综合性立法尝试,其人工智能条款(Subtitle C, Part Ⅱ)集中体现了这一矛盾本质。该法案第 43201 条主要通过减少监管负担促进人工智能创新,但实质通过三项机制固化国内监管真空:其一,非强制性框架主导。法案要求 NIST 制定“自愿性人工智能风险管理框架”,明确禁止将其转化为强制性标准,使企业合规完全依赖自律。其二,象征性协调机制。设立“人工智能跨部门工作组”,但其职能仅限于“研究监管差距”和“提出非约束性建议”,无权制定统一规则。其三,创新豁免特权。允许企业通过“安全港”条款规避州级算法透明度诉讼,实质削弱消费者救济渠道。该条的第(c)款标题则是直接叫作“中止令”(Moratorium),禁止各州及其下属政治机构在十年内通过立法来限制人工智能技术和产业的发展。

这种立法缺位绝非偶然,而是维护技术霸权的战略选择。H.R.1 法案第 305 条要求美国商务部评估“人工智能出口管制对创新的影响”,却对国内算法歧视、隐私侵犯等风险只字未提,揭示其核心关切在于技术竞争而非权利保障。更关键的是,法案将国防、情报领域人工智能应用完全排除在监管范围外[§307(c)],延续《芯片与科学法案》第 10302 条确立的“国家安全例外”原则,[34]

---

〔29〕 The Office of Science and Technology Policy (OSTP), The Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People, The White House, October 2022, https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

〔30〕 National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, 2023.

〔31〕 参见郑戈:《数字主权视角下国家对人工智能的监管框架比较》,第 35 页,未刊稿。

〔32〕 NIST AI RMF 1.0, p.ⅲ.

〔33〕 Executive Order No. 14110, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 88 Fed. Reg. 75191 (2023).

〔34〕 CHIPS and Science Act, Pub. L. 117-167 §10302 (2022).

使五角大楼和情报机构得以无限制开发进攻性AI工具。这种“内松外紧”的规制二元结构，印证了美国模式的本质——国内市场的自由化表象建立在对外技术封锁的基石之上。

联邦层面的消极立法直接导致监管碎片化。H.R.1法案第308条禁止州政府制定严于联邦标准的人工智能规则，但联邦自身未设定任何实质性义务，形成“无标准即无限制”的悖论。其结果如康涅狄格州《人工智能权利法案》被迫删除算法审计条款，纽约市《自动化就业决策工具法》因联邦不行动而陷入执法困境。这种制度设计迫使各州监管沦为“无牙之虎”，正如加州隐私保护局2025年度报告所指：“联邦立法真空使州级权利保障机制失去锚点。”〔35〕

美国模式的深层逻辑在于：通过规避国内监管成本维持技术代差优势。H.R.1法案草案第43201条授权拨款20亿美元资助AI研发，却要求资金仅用于“突破性技术”而非合规建设。这种资源分配折射出战略优先级——技术创新服务于国家安全目标，而非公民权利保障。当法案将“人工智能领先地位”定义为“核心国家安全利益”，其国内不立法的实质已昭然若揭：对内放松监管以鼓励人工智能技术和产业发展，对外安全至上以遏制竞争对手的发展。

2. 泛安全化的联邦人工智能相关立法

美国虽在国内治理中高举“市场自由”旗帜，但其“涉外法治”始终以技术遏制为核心。2022年《芯片与科学法案》以“国家安全”名义禁止受补贴企业在华扩建先进制程产能，2024年《保护美国人数据免受外国敌对势力侵害法案》强制TikTok剥离业务，联邦最高法院在TikTok案〔36〕中更以“国家安全例外原则”为行政扩权背书。这种“内松外紧”的二元结构暴露美国模式的实质：国内不立法是为降低企业合规成本，加速技术迭代以维持全球领先；对外激进立法则是为消除战略竞争对手的威胁。国家人工智能安全委员会（NSCAI）在2021年报告中直言不讳地将中国定位为“最大竞争对手”，主张通过出口管制与人才流动控制遏制中国AI发展。

因此，“市场主导”仅是表象，其本质是以国家安全为轴心的技术霸权护持模式。当阿努·布拉德福德将美国归类为“市场驱动型”时，〔37〕她忽视了市场繁荣与联邦立法缺位的因果关系：正是国内监管的刻意留白，为科技巨头创造了全球扩张的制度优势。但这种模式已陷入自我消解的悖论：为压制中国技术崛起而强化的对外管制（如AI芯片出口禁令），反而迫使美国企业依赖宽松的国内环境加速创新，导致公民权利保障持续让位于战略竞争需求。特斯拉自动驾驶事故中，企业以“商业秘密”抗辩拒绝公开算法代码，受害者难以获取关键证据，正是这一悖论的现实注脚。

与针对国内市场的立法缺位相比，当技术发展触及国家安全红线时，联邦权力便以防御性姿态强势介入。近年立法实践清晰表明，安全关切已完全主导美国人工智能规制路径：2022年《芯片与科学法案》以国家安全为由禁止先进芯片技术对华输出，构建半导体产业“小院高墙”；2024年《保护美国人数据免受外国敌对势力侵害法案》强制字节跳动剥离TikTok，开创以数据主权名义限制特定外国平台的先例；联邦最高法院在TikTok案中更以国家安全例外原则为上述措施背书，确立行政机构在数字领域的超宽裁量权。此类立法呈现高度选择性特征：其规制对象聚焦中国相关实体，规制手段依赖出口管制、投资审查等单边工具，规制目标直指技术竞争而非风险治理。

这种“内松外紧”的立法二元结构暴露美国模式的实质：国内市场的自由化表象建立在对外系统性规制的基石之上。通过将中国人工智能发展建构为国家安全威胁，美国得以正当化其背离市场原则的干预措施。芯片管制削弱竞争对手技术基础，TikTok禁令阻断数据跨境流动，司法判决

〔35〕 California Privacy Protection Agency (CPPA), *Annual Report on AI Governance*, 2025, p.23.

〔36〕 TikTok *v.* Garland, 145 S. Ct. 57, 2025.

〔37〕 Anu Bradford, *supra* note〔1〕, at 73.

则为安全例外提供宪法容器，三者共同构成“以安全为名、以遏制为实”的规制闭环。其后果是形成功能性悖论——对外安全立法越强化，越需维持国内监管宽松以加速技术迭代，最终导致国内用户权利保障与算法问责持续让位于战略竞争需求。这种安全至上的单边主义路径，不仅与欧盟风险治理框架形成鲜明对照，更凸显布拉德福德模型对立法动机与制度功能的结构性误判。〔38〕

综上所述，将美国人工智能立法模式简单归类为“市场主导”，未能穿透其市场表象触及立法行动的核心驱动力。美国国内市场的高度活力是历史与制度形成的既定事实，而非其立法者在国内人工智能治理上的主动模式选择。就立法行动本身而言，美国联邦层面的核心特征是对内监管的战略性谦抑与对外（主要对华）技术管制的国家安全立法激进化。前者旨在避免束缚创新手脚，后者则旨在消除战略竞争对手的威胁。两者相辅相成，共同服务于维护美国全球科技霸权与国家安全的最高目标。因此，“安全至上”才是理解美国人工智能立法模式实质的钥匙，它超越了简单的“国家—市场”二分法，揭示了地缘政治竞争深刻重塑技术治理格局的时代特征。忽略这一本质，任何对美国人工智能立法模式的讨论都将失之偏颇。

五、结　　论

人工智能立法的全球实践表明，立法模式差异的本质在于价值预设的深层分歧，这种分歧通过立法目标、工具选择与实施效力的差异化配置形成制度表达。中国的“有序创新”模式以技术发展与风险控制的动态平衡为核心，通过渐进式立法、地方试验性规则与阶段性规划构建弹性治理框架，其制度内核在于以创新效能最大化对冲系统性风险。典型如《深圳人工智能产业促进条例》设立专项基金支持算法研发，同时建立伦理委员会提供合规指导，形成公私协同的创新激励机制。美国的“安全至上”模式呈现为对内监管谦抑与对外规制强化的二元结构：联邦层面通过《通信规范法》第230条豁免平台责任，维持数据自由流通以降低本土企业合规成本；涉外领域则依托《芯片与科学法案》实施技术封锁，以国家安全例外机制维系技术霸权。欧盟的“风险规避”模式植根于产品安全监管范式移植，通过四层风险分级将人工智能系统定性为“危险产品”，依托 TFEU 第114条统一内部市场规则，基本权利保护仅作为次要目标嵌入合规要求（如 AIA 第15条强制日志记录机制）。

由此可提炼出三项一般性命题：其一，立法价值取向构成制度设计的元规范，直接决定法律规则的风险识别精度与责任分配逻辑。当立法者将“国家安全”设定为价值优选项时（如美国），必然催生以技术遏制为导向的涉外法治体系；当“风险规避”成为核心关切（如欧盟），则必然形成预防性合规义务的刚性约束；而“有序创新”的价值定位（如中国）必然要求构建容错机制与产业促进政策的制度组合。其二，价值预设通过工具理性实现对象化，立法工具的选择实质是价值冲突的技术化解决方案。欧盟将伦理原则转化为可验证的技术标准（如 HLEG《可信人工智能伦理指南》要求高风险系统设立人工干预机制），本质是以工程语言重构权利话语；中国通过《数据安全法》首创数据分类分级制度，实则是以数据要素管控实现创新秩序建构；美国依赖行业自律与州级分散立法，反映市场优先价值对统一规制的拒斥。其三，全球人工智能治理的本质是价值共识的竞争性建构过程，国家立法差异实际上是不同文明传统对“人机关系”的规范性回应。欧盟强调“人在回路”原则（如 AIA 第14条要求高风险系统确保人类监督），中国注重“科技向善”的伦理内嵌（如《生

〔38〕 Aziz Z. Huq, *The Geopolitics of Digital Regulation*, 92 The University of Chicago Law Review 833 (2025).

成式人工智能服务管理办法》要求训练数据符合核心价值观），美国侧重算法结果控制（如纽约市 Local Law 144 规范招聘算法偏见），三者共同指向“人类主导—AI 辅助”的底层共识，差异仅在于实现路径的制度化程度。

当前全球立法实践显示出两项结构性约束：技术迭代速度与法律滞后性之间存在永恒张力，过度依赖预设风险清单（如 AIA 附件Ⅲ）将导致规则失焦，需通过动态授权立法（如中国网信办持续更新算法推荐管理规定）保持制度适应性；主权边界与技术全球化存在根本冲突，美国外资审查机制与欧盟“布鲁塞尔效应”的博弈表明，跨境数据流动规则实质是物理空间的领土边界向数字空间的延伸。未来立法需在三个维度实现突破：建立价值取向的弹性转换机制，允许特定场景下安全本位向创新导向的合理让渡（如中国自动驾驶地方立法中的容错条款）；发展“监管沙盒”与“合规宽恕”等弹性制度，通过试错成本社会化降低创新阻力；构建跨司法辖区的互认框架，在生物识别、医疗诊断等高风险领域率先形成监管共识，避免技术标准碎片化割裂全球市场。最终，人工智能立法的演进方向应是价值理性与工具理性的辩证统一——既承认国家价值选择的正当性差异，亦通过技术治理的共通语言弥合制度鸿沟。

Abstract　The current popular classification of global artificial intelligence legislative models, such as the state-led model represented by China, the market-led model typified by the US, and the EU's rights-oriented model, is essentially a labeling exercise based on ideological positions. It fails to accurately reflect the core driving forces and practical reason of AI legislations in different countries. This paper posits that understanding the differences in AI legislation should focus on intrinsic value orientations. China's legislative practice centers on "ordered innovation", aiming to balance technological development with risk control, and enhance industrial competitiveness through phased planning. The US prioritizes "national security" in its legislation, with regulation logic mainly serving to maintain technological hegemony and address geopolitical challenges. The EU, on the other hand, demonstrates a clear "risk-averse" approach, leveraging strict ex-ante and preventive rules (like the GDPR and the AI Act framework's detailed compliance requirements) to minimize the social risks of technological application. This value-based classification better reveals the underlying causes of legislative differences and offers a more objective analytical framework for understanding the global AI governance landscape.

Keywords　Artificial Intelligence Legislation, Value Orientation, Ordered Innovation, Risk Aversion, National Security

（责任编辑：徐舒浩）