

欧盟寻求人工智能监管规则型领导地位： 认知与路径^{*}

□ 邢亚杰 李丹成

〔提 要〕在数字时代，主要大国或国家联盟围绕技术型与规则型两类领导地位展开竞争。基于维护数字主权、防范竞争对手以及评估自身优劣势的综合认知，欧盟将确立监管规则型领导地位视为自身核心战略目标，成为创设监管规则的先行者。为打造人工智能监管的示范版图，欧盟不断宣扬人本伦理的作用，在实际操作中赋予人本伦理以能动性与约束力，进而构建了软硬结合的监管框架，意图打造技术风险等级分明、配套监管措施完备、风险防范和技术创新平衡的监管规则体系。欧盟的尝试是理想与现实妥协的产物，有助于提升其在人工智能监管领域建章立制的影响力，在一定程度上赋能全球人工智能治理。但欧盟的做法加剧了域内外企业的负担，激化了全球人工智能监管规则之争。由于国际大格局束缚、各方竞争加剧、全球人工智能治理结构性挑战层出不穷、欧盟政策有缺陷等原因，欧盟的诉求存在较大不确定性。

〔关键词〕人工智能、欧盟、政府监管、数字竞争

〔作者简介〕邢亚杰，清华大学社会科学学院助理研究员

李丹成，清华大学公共管理学院博士研究生

〔中图分类号〕TP18, F495

〔文献标识码〕A

〔文章编号〕0452 8832 (2025) 2 期 0108-21

^{*} 本文为教育部人文社会科学重点研究基地重大项目“百年大变局下中国与中东欧国家（16+1）合作机制发展的现状、挑战及前景研究”（编号：22JJD810030）的阶段性成果。

人类社会数字化发展趋势明显，数字科技对各国、各区域经济发展与公共治理意义重大，世界主要政治经济大国或国家联盟都已充分意识到数字科技的革命性力量，将其提升到关乎自身竞争力与安全的战略高度。人工智能作为新兴交叉科技，既致力于开发数字设备的类人思考与行为能力，也关注如何充分挖掘各数字要素的价值，因而被认为是数字科技皇冠上的明珠，对国家竞争力提升具有革命性意义。多年来，全球科技界深耕移动互联网、大数据、超级计算等相关领域研究，为人工智能的厚积薄发提供了重要基础。2022 年底以来，ChatGPT、DeepSeek、Manus 等生成式人工智能成果相继发布，在世界范围内引发强烈反响。人工智能研发迅猛提速，呈现出跨界融合、人机协同、群智开放等特征，展现出巨大潜能。

然而，信息伪造、网络盗窃、知识产权侵害、认知偏见、个人隐私泄露、失业风险等一系列与人工智能相关的新问题、新挑战随之激增。世界各地政府对此高度关注，纷纷探求有效监管之道，其中欧盟的表现尤为突出。早在 2021 年 4 月，欧盟就“先知先觉”，对外公布《人工智能法案》草案，旨在出台全球首个关于人工智能监管规则与标准的法案。2023 年 6 月，欧洲议会以压倒性优势表决通过了《人工智能法案》授权草案，这也是全球第一个正式落地的人工智能监管横向立法。2023 年 12 月，欧洲议会、欧盟成员国和欧盟委员会达成正式实施该法的相关临时协议。^[1]2024 年 8 月，《人工智能法案》正式生效。

人工智能监管正在成为全球数字竞争的新高地，国内外高度关注美国与中国的相关动向，^[2]对欧盟的研究却显不足。研究欧盟追求人工智能监管规

[1] European Parliament Press Center, “Artificial Intelligence Act: Deal on Comprehensive Rules for Trust Worthy AI,” December 9, 2023, <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>.

[2] 参见黄日涵、姚浩龙：《被重塑的世界？ChatGPT 崛起下人工智能与国家安全新特征》，《国际安全研究》2023 年第 4 期，第 82-106 页；戚凯：《ChatGPT 与数字时代的国际竞争》，《国际论坛》2023 年第 4 期，第 3-23 页；孙学峰：《数字技术创新与国际战略竞争》，《外交评论》2023 年第 1 期，第 54-77 页；余南平：《新一代通用人工智能对国际关系的影响探究》，《国际问题研究》2023 年第 4 期，第 79-96 页。

则型领导地位的认知、成因与举措，并对其影响与限度进行评估，有助于实现有效的全球人工智能治理。

一、全球数字领导地位与人工智能竞争新场域

数字技术的崛起使全球政治和经济秩序发生根本性的变革，大国或国家联盟之间的竞争不再局限于传统现实空间，还扩展到了数字平行世界，这不仅是一场数字技术的较量，更是关系到自身安全、经济发展与社会稳定的复合型竞争。在这场复合型竞争中，一些大国或国家联盟为了树立在数字空间中的领导地位，不仅大力促进技术创新，也极力寻求规则与理念上的领导地位。

地位是解释国家或国家联盟行为的重要变量，它被广泛地理解为国家或国家联盟在国际等级体系中的位置或成员资格。综合来看，目前关于地位的观点主要集中在以下几方面：第一，地位是一个主体间概念，国家或国家联盟通过彼此间的比较确立自身的地位；第二，国家或国家联盟在国际等级体系中的位置由自身的硬实力和软实力决定；第三，地位具有变化性，是许多相互交叉和重叠的行动综合作用的结果，地位可被争夺、塑造与再塑造；第四，国家或国家联盟关注自身在世界上的地位，出于国际体系中的竞争压力、平衡国内利益诉求的考量，会维持既有优势地位或寻求地位的跃升；第五，地位是一种稀缺性资源，所有参与方不可能拥有同等的地位，这是产生地位竞争的原因；第六，中观层面的地位理论主要以地位寻求为分析原点，形成地位认知、地位策略与地位承认的分析链条。^[1]

[1] Ryan McNamara, "The Status of Status in International Relations," *MA Major Research Papers*, 2022, p.16; Jonathan Renshon, "Status Deficits and War," *International Organization*, Vol.70, No.3, 2016, pp.513-550; Kai He and Huiyun Feng, "Role Status and Status-saving Behaviour in World Politics: the ASEAN Case," *International Affairs*, Vol.98, No.2, 2022, pp.363-381; 王梓元：《国际政治中的地位与声望：一项研究议程》，《国际政治研究》2021年第3期，第116-139页；刘博文：《国际承认、国内政治与大国的地位追求策略》，《世界经济与政治》2022年第5期，第101-130页；孙志强、张蕴岭：《地位寻求、角色构建与英国脱欧后的对外政策调整》，《欧洲研究》2022年第4期，第106-134页。

传统现实空间中的国际权力格局与地位态势存续多年，已成为国际社会某种意义上的既定现实。与此同时，随着人类社会发展内容与数字议题深度嵌合，数字空间与传统空间正在不断融合，有学者指出，与“领土/空/海”对应的“领云”（clouds territory）这一全新疆域空间正在逐步成形，它是国家或国家联盟地位寻求的全新赛场。^[1] 美西方高度重视自身在全新空间的地位水平，积极寻求领导地位的平移。从这个角度出发，全球数字竞争的本质是大国或国家联盟在数字空间中领导地位的竞争，诸如美国这样的传统霸权国高度关注自身在数字空间的霸权构建能力，意图在“领云”复制唯一领导者地位。

全球数字竞争过程中，大国或国家联盟寻求的领导地位类型主要有两种：第一种是技术型领导地位，处于这类地位的大国或国家联盟在数字技术创新和研发投入、人才储备和培养以及政策支持上有领先优势；第二种是规则型领导地位，处于这类地位的大国或国家联盟在数字经济、网络安全等方面拥有较强的法规制定和执行能力，能够制定并推动行业性或全球性标准。大国或国家联盟在数字空间中寻求领导地位时，会根据自身实际情况重点寻求最能实现的领导地位，并且不拒斥同时寻求两种领导地位。

一方面，技术型领导地位的竞争趋向深化。研发投入是反映创新能力的核心指标，欧盟委员会报告显示，全球各地的数字科技研发投入都显著增加，占全球各行业总研发投入 40% 以上。^[2] 除此之外，美西方高度关注人才这一推动技术创新的内生动力，特朗普上任后迅速成立“总统科学技术顾问委员会”，旨在团结各界高端人才实现并保持美国的全球技术领导地位；欧盟发布《2023—2024 年数字欧洲工作计划》，特别强调要提升吸引培育数字人才的能力，营造有利于创新和数字化转型的人才系统，持续推进“深层技术人才倡议”。

[1] Benjamin H. Bratton, *The Stack: On Software and Sovereignty*, The MIT Press, 2015.

[2] European Commission, “Industrial Investments in Research and Development in the EU again on the Rise,” December 13, 2023, https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7647.

另一方面，规则型领导地位的竞争日益加速。欧盟出台《人工智能责任指令》、《人工智能法案》等多份战略性文件，力图将欧盟的规则标准上升为全球的规则标准。英国持续发挥在制定经合组织人工智能原则上的主导作用，希望以建立国际联盟、加强多边合作的方式增加英国在全球数字技术标准机构中的代表性与领导权。印度也一改鼓励人工智能创新发展的常态，强调人工智能监管势在必行，印度的电信监管局、标准局及储备银行积极制定相关监管框架或规则。^[1]

另外，值得注意的是，美西方越发注重将价值观要素融入技术型与规则型领导地位的内容中。英国政府通讯总部提出，民主、法治、人权等价值观是社会的基础，技术研发与应用必须符合这些价值观。^[2] 传统基金会科技政策中心总监卡拉·弗雷德里克预测，特朗普政府未来人工智能政策的重心之一，就是确保美国人工智能技术符合西方的自由民主价值观。^[3]

ChatGPT 的走红促使人工智能在全球科技、金融、政府等各界迅猛发展，行业发展迎来前所未有的超级红利期，人工智能由感知智能向认知智能、由专用智能向通用智能的跃迁趋势越来越明显。国际社会普遍认为，以此轮人工智能跃迁为关键契机，人类数字化发展的水平全面提速，人类数字空间的质变为期不远。与之伴随的则是，大国或国家联盟对全球数字领导地位的竞争迎来人工智能赛场中的“狂飙”时刻。

其一，投资明显升温。为了赢得生成式人工智能竞赛，美欧政府和企业掀起新一轮投资热潮。2025 年 1 月，英国出台《人工智能机遇行动计划》，强调加大计算资源投资，并承诺提供 10 年的投资支持。^[4] 随后，特朗普宣布

[1] “Regulating AI: Navigating India’s Challenging Regime,” April 15, 2024, <https://law.asia/navigating-ai-india/>.

[2] Government Communication Service, “GCS Framework for Ethical Innovation,” September 2021, pp.2-3, https://gcs.civilservice.gov.uk/wp-content/uploads/2024/05/2024-04-GCS-Framework-for-Ethical-Innovation_GCS-template-v1.3-OFF-SEN.pptx.pdf.

[3] Center for Strategic and International Studies, “AI Policy in Trump 2.0,” December 16, 2024, <https://www.csis.org/podcasts/ai-policy-podcast/ai-policy-trump-20>.

[4] “AI Opportunities Action Plan,” January 13, 2025, <https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan>.

启动由软银集团、开放人工智能研究中心及甲骨文共同参与的“星际之门”人工智能基础设施计划，预计投入 5000 亿美元建设数据中心、计算设施等关键技术平台。^[1]2 月，欧盟宣布启动“投资人工智能”倡议，拟调动 2000 亿欧元投资，以推动人工智能发展。^[2]这些政府与企业的投资规模已远超高盛对 2025 年全球人工智能投资额达到约 2000 亿美元的预测。^[3]

其二，规则领导权之争更加激烈。对人工智能严加监管成为一些大国或国家联盟打造政绩、塑造全球领导力的着力点。欧盟加速人工智能立法，自 2023 年中期以来，《人工智能法案》推进速度明显提升；新加坡先后发布《生成式人工智能治理模型框架》《人工智能系统安全指南》，创立非约束性监管模式；韩国迅速出台《人工智能发展和建立信任基本法》，成为全球第二部人工智能立法；西方传统大国在英美澳同盟、七国集团、经合组织、北约等联盟框架内部的人工智能规则构建行动更是层出不穷。

其三，价值观对抗色彩加重。美国等大国或国家联盟秉持“己优他劣”“非此即彼”的二元对立认知，将意识形态对抗、价值观之争全力融入此轮人工智能发展竞赛中。2025 年 2 月，美国副总统杰德·万斯在人工智能行动峰会上化身“两面派”，时而宣称人工智能必须保持不受意识形态偏见的影响，时而刻意反对“独裁政权”使用人工智能，^[4]宣扬意识形态对立，意图推进科学技术政治化。

[1] 《特朗普开启“星际之门”：21 世纪 AI 时代的“星球大战”？》，央视网，2025 年 1 月 26 日，<https://news.cctv.com/2025/01/26/ARTIgx6BpURgflLxcaWkVtbl250126.shtml>。

[2] 《欧盟拟以“投资人工智能”倡议调动 2000 亿欧元》，新华网，2025 年 2 月 11 日，<http://www.news.cn/world/20250211/1502db6563634e1db8b4d01648bf0e73/c.html>。

[3] Will Canny, “AI Investment Forecast to Approach \$200 Billion Globally by 2025,” August 3, 2023, <https://www.goldmansachs.com/intelligence/pages/ai-investment-forecast-to-approach-200-billion-globally-by-2025.html>。

[4] The American Presidency Project, “Remarks by the Vice President at the Artificial Intelligence Action Summit in Paris, France,” February 11, 2025, <https://www.presidency.ucsb.edu/documents/remarks-the-vice-president-the-artificial-intelligence-action-summit-paris-france>。

二、欧盟寻求人工智能监管规则型领导地位的认知

随着数字技术崛起与人工智能跃迁，人工智能监管与数字领导地位密切相关，不少强势竞争者都在积极寻求人工智能监管规则的主导权。譬如，英国成立人工智能基础模型工作组，主办首届全球人工智能安全峰会，旨在成为“全球人工智能安全监管的地理中心”。^[1]受多方因素影响，欧盟在这场领导地位的竞争中确立了寻求人工智能监管规则型领导地位的战略目标。

（一）维护数字主权的地位认知

欧盟将数字竞争视为全球地缘政治博弈的核心内容之一，将人工智能等尖端数字科技视为影响国际权力再分配和国际关系再塑造的重要力量。欧盟认识到，数字科技的迅猛发展会带来一系列安全与发展难题，一旦错失机遇或竞争失利，本区域的公民、企业和成员国政府会逐渐失去对数据、创新能力以及在数字空间中立法执法能力的控制。因此，维护欧盟数字主权是其数字外交战略的底层逻辑，与此同时，欧盟也渴望增强自身在数字空间中的领导力和战略自主权，并成为全球技术地缘政治的主要玩家之一。^[2]

在欧盟委员会主席冯德莱恩的推动下，数据安全、数字主权理念已经成为欧洲参与数字和技术事务、提升数字竞争力和话语权的指导原则。欧盟在数据保护领域首创《通用数据保护条例》，在维护数字主权问题上开创性地提出了“致力于支持欧盟内部的技术发展、制定管理数据和数字环境的全球规范、对欧盟市场上的非欧盟参与者施加更多限制”三大原则，上述举措在

[1] Ryan Browne, “British Prime Minister Rishi Sunak Pitches UK as Home of A.I. Safety Regulation as London Bids to Be Next Silicon Valley,” June 12, 2023, <https://www.cnn.com/2023/06/12/pm-rishi-sunak-pitches-uk-as-geographical-home-of-ai-regulation.html>.

[2] Julian Ringhof and José Ignacio Torreblanca, “The Geopolitics of Technology: How the EU Can Become a Global Player,” May 17, 2022, <https://ecfr.eu/publication/the-geopolitics-of-technology-how-the-eu-can-become-a-global-player/#the-toolbox-for-eu-digital-diplomacy>.

全球范围内引起广泛关注。^[1]

规则型领导地位的寻求是欧盟维护数字主权的核心目标和主要路径。欧盟明确提出“要实现数字主权，欧盟必须建立一个真正的数字单一市场，加强制定规则的能力，自主作出技术选择……在国际层面，欧盟将利用其工具和监管权力，帮助制定全球规则和标准”。^[2]这种诉求表达意味着欧盟对规则型领导地位的认知是内外兼顾的，欧盟想要形成所谓的规则型领导地位，既要有内部的市场环境与自主能力的联合支撑，也要有将内部监管实践与经验转化为对外软实力的能力，从而引领全球人工智能监管规则与标准的制定。

进一步说，欧盟在维护数字主权、增强数字竞争力、寻求人工智能监管规则型领导地位之间，构建了一套环环相扣的认知体系：其一，采取更加积极的数字经济刺激政策，扩大欧盟数字经济规模与影响力；其二，以人工智能产业为数字经济扩张目标的主要抓手，通过对人工智能领域的更多投资来催生本地领军企业；其三，通过高标准的人工智能监管，提升对域外人工智能企业巨头的拒止能力，提高自身数字安全水平；其四，依托一系列人工智能治理建章立制，塑造欧盟规范先进的治理者形象。各项措施多管齐下，最终合力确立欧盟的人工智能监管规则型领导地位，实现规则外溢，提振欧盟数字竞争力。

（二）对竞争对手优势的威胁性认知

从资金投入、技术发展、人才培养、应用场景等多方面来看，中美两国已显著占据全球人工智能领域“两极”之位。^[3]欧盟对中美两国形成了较强烈的威胁性认知：第一，对中美人工智能技术大幅领先的高度忧虑。在本轮

[1] Frances Burwell and Kenneth Propp, “Digital Sovereignty in Practice: The EU’s Push to Shape the New Global Economy,” November 2, 2022, https://www.atlanticcouncil.org/wp-content/uploads/2022/11/Digital-sovereignty-in-practice-The-EUs-push-to-shape-the-new-global-economy_.pdf.

[2] European Council, “Special Meeting of the European Council (1 and 2 October 2020)-Conclusions,” October 2, 2020, <https://www.consilium.europa.eu/media/45910/021020-euco-final-conclusions.pdf>.

[3] 施雯、缪其浩：《从两极到三强：欧盟人工智能的全球竞争战略分析》，《中国科技论坛》2022年第6期，第180页。

生成式人工智能爆火的过程中，从技术水平与产业规模来看，中美两国的确较大幅度领先于欧盟。美国拥有全球最多最大的生成式人工智能公司，这些公司在欧洲占有绝大部分市场份额。^[1] 美国还拥有大规模数据集和强大的计算资源，以及各种开源软件库与机器学习库，形成了领先的人工智能模型训练和优化能力。中国的表现也较为抢眼，2024 年新增的 45000 条生成式人工智能专利中，中国申请数量高达 27000 条，占比 61.5%，位列全球第一。^[2] 中国人工智能企业“深度求索”推出性能优越、免费商用的开源大模型，且训练成本相较同类产品更低，在国际上引起广泛关注和热烈讨论。目前，欧盟几乎所有的人工智能前沿业务都与美国或中国有关联，大多数欧洲企业使用两国公司提供的应用程序接口，且欧盟自有的通用式人工智能系统的培育尚未成功，与国际竞争对手的差距较为明显。这种过度依赖的局面加剧欧盟的安全焦虑。

第二，对美国开发人工智能敏感应用的高度紧张。美国一贯对人工智能军事化开发、人工智能情报搜集、人工智能“混合战争”等领域有强烈兴趣。欧盟绝大多数国家均在北约框架下受到美国军事霸权的控制，又曾多次遭遇“棱镜门”等美方网络黑客行动侵害，因此欧盟极其忧虑美国不断深化人工智能科技在军事层面的运用，将使其完全处于下风，国防与情报等底线安全毫无保障。与此同时，出于根深蒂固的意识形态畸见，加之美国煽动，欧盟也猜忌中国的人工智能发展，认为中国会利用人工智能科技对其进行所谓的“自由压制”，侵害其基本人权与隐私。

第三，对中英积极布局人工智能监管的担忧。出于对人类命运共同体的高度关切，中国亦高度重视人工智能风险及治理，近些年相继发布了《全球人工智能治理倡议》《人工智能安全治理框架》等指导性文件，提出了“以

[1] “The Geopolitics of Generative AI: International Implications and the Role of the European Union,” Elcano Royal Institute, November 17, 2023, <https://www.realinstitutoelcano.org/en/work-document/the-geopolitics-of-generative-ai-international-implications-and-the-role-of-the-european-union/>.

[2] 《2024 年生成式 AI 专利调查：中国新增量全球第一》，澎湃新闻，2025 年 2 月 18 日，https://www.thepaper.cn/newsDetail_forward_30100207。

人为本”“智能向善”的适度监管理念，围绕分类分级保护、安全审查、安全管理等进行了若干制度设计，在国际社会特别是“全球南方”产生了热烈反响。英国则致力于推广“英式方案”，形成根据不同行业实施不同监管规定的“第三条道路”。^[1]中英的这些做法，毫无疑问都会对欧盟领行全球人工智能监管的战略目标造成挑战，因此也正在受到欧盟的密切关注，并刺激其更加焦虑。

（三）对欧盟自身优劣势的认知

欧盟对其在数字科技，特别是人工智能领域的优势与劣势具有若干认知，其中一些方面是较理性与清晰的，在另一些方面却又时常因为“有色眼镜”而带有较强的感情色彩。

首先，欧盟是技术层面的“落后者”。尽管欧洲取得了一定的数字科技进步，但整体而言，欧洲在数字化和人工智能领域滞后，欧盟对此有着清晰的理性认知。相较于中美两国，欧盟的科技创新投资不足，公共和私人投资的缺乏阻碍了生成式人工智能技术的发展，创新生态系统也较为分散，缺乏一体化和协同创新的优势。欧盟企业在人工智能等前沿数字技术的应用方面相对滞后，2023年，仅有8%的企业使用人工智能技术，2024年这一比例虽有所提升，但仅达到13.48%。^[2]欧盟非常清楚，技术研发水平与应用动力不足是自身的硬伤，这使得欧盟难以深度参与人工智能技术型领导地位争夺，寻求规则型领导地位是欧盟唯一可期待的目标。

其次，欧盟是监管层面的“领先者”。在数字科技监管领域，欧盟有着丰富的立法尝试，欧盟被普遍认为是全球监管最严厉的地区，其早在2004年就通过了第一版《数字服务法》，这是全球首部对数字化市场、产品和服务

[1] “The UK Wants to Do Its ‘Own Thing’ on AI Regulation, Suggesting a Divergence from U.S. and EU,” NBC New York, January 14, 2025, <https://www.nbcnewyork.com/news/business/money-report/the-uk-wants-to-do-its-own-thing-on-ai-regulation-suggesting-a-divergence-from-u-s-and-eu/6105957/>.

[2] European Commission, “Use of Artificial Intelligence in Enterprises,” January 2025, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use_of_artificial_intelligence_in_enterprises.

实施立法与依法监管的法律；2018 年通过的《通用数据保护条例》，成为全球数据隐私保护的标杆，也被视为隐私保护领域“史上最严”立法。为适应数字技术的发展，2020 年 2 月以来，欧盟累计制定了 50 余项综合性文件，包括法案政令、产业与安全战略、合作指南、临时政治协议、政策建议等。这些前期积累使得欧盟在全球同侪中显得富有经验，而人工智能监管领域恰好缺乏统一和明确的全球性框架，欧盟因此认为自身寻求规则型领导地位可谓“厚积薄发、舍我其谁”。

最后，欧盟是道德层面的“居高临下者”。作为现代国际关系体系与西方文明思想的发源地，欧洲一直将自己视为所谓“普世文明”的代言人和守护者，这种自傲感使得欧盟国家特别注重个体本位的隐私保护、数据安全、人权正义，并且惯于对其他国家或地区的自主选择指手画脚。受道德伦理层面的优越感影响，欧盟对人工智能监管事业有着外界难以理解的“高尚使命感”和“神圣责任感”，其处处宣称要对人工智能技术创新实现指导与约束，建立以人为本的数字治理愿景。正如欧盟委员会执行副主席兼竞争专员玛格丽特·维斯塔格（Margrethe Vestager）所言：“我们相信以人为本的数字化转型是一场没有人被抛下的转变。我们想要为人们服务的安全技术。”^[1]

总的来看，欧盟决意寻求人工智能监管规则型领导地位的认知是内外因素综合作用的结果，既有维护数字主权的目的，也为了对冲中美人工智能发展带来的压力，同时还有扬长避短的考虑。

三、欧盟寻求人工智能监管规则型领导地位的路径

欧盟试图基于自身的比较优势寻求人工智能监管规则型领导地位，并通过以下路径构建人工智能监管的示范版图，致力于成为此轮全球数字竞争中

[1] European Commission, “Speech by Executive Vice-President Vestager on the Declaration on Digital Rights and Principles,” January 26, 2022, https://ec.europa.eu/commission/presscorner/detail/en/speech_22_625.

的核心玩家与人工智能领域的“第三极”。^[1]

（一）强调“以人为本”的伦理基础

欧盟拥有重视个人隐私及数据相关权利的价值传统与民意基础，立足自身经验，精准把握西方世界“价值观外交”回潮的时代趋势，提出了“以人为本”的伦理立场，抢先发布人工智能的伦理监管准则并将其推向国际社会，尝试以欧洲方式实现“弯道超车”。近年来，欧盟已经密集出台了《欧洲人工智能战略》《人工智能协调计划》《人工智能白皮书》《人工智能法案》等一系列具有重要意义的政策文件。欧盟极力宣扬“以人为本”的伦理基础，并将其视为人工智能监管规则的支撑理念和政策基石。^[2] 欧盟通过以下两个步骤明确“以人为本”基础性地位：第一，宣扬“以人为本”的规范性力量，使人工智能监管规则成为塑造可信赖、安全人工智能技术的要素。欧盟以应对人工智能开发与应用中的不透明性、歧视与偏见泛滥、激化犯罪、侵犯隐私等风险为主要理由，将人本主义、透明度、公平、人权等普遍性伦理要素内化为“以人为本”的核心构成，并进行规范性力量的投射。例如，《人工智能白皮书》明确指出，欧洲的人工智能技术应以人为中心、符合道德与可持续发展、尊重人类基本权利和欧洲价值观；^[3]《人工智能法案》致力于促进“以人为本”和值得信赖的人工智能的使用，民主与人权等价值观再次成为技术规制的核心规范。^[4]

第二，注重“以人为本”在领导地位竞争中的战略嵌入。欧盟强调，其

[1] Cezara Panait, Davor Ljubenkov and Dalia Alic, “Striking the Balance between Innovation and Regulation in AI—Is Europe Leading the Way or Lagging Behind?,” *Europuls Policy Journal*, Vol.1, No.1, 2021, pp.27-45.

[2] European Commission, “Communication: Building Trust in Human Centric Artificial Intelligence,” April 8, 2019, <https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence>.

[3] European Commission, “White Paper on Artificial Intelligence: A European Approach to Excellence and Trust,” February 9, 2020, p.2, https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

[4] European Parliament, “Artificial Intelligence Act,” June 2023, p.2, [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf).

人工智能监管规则的不同之处在于：欧盟致力于在保护并尊重人类基本权利和价值观的前提下改善公民生活，助力社会 and 经济发展，进而使欧盟及其成员国成为人工智能“以人为本”发展道路上的全球领导者。此定位透露出欧盟将伦理原则嵌入领导地位竞争的内在逻辑：首先，“以人为本”的人工智能必须符合欧盟伦理，即欧盟人权宪章级文件《里斯本条约》中明确提到的“必须建立在尊重人的尊严、自由、民主、平等、法治和尊重人权，包括少数族裔人权的基础上”。这些伦理要素应该完全融入人工智能技术与系统从开发到应用的全过程。其次，符合欧盟伦理的人工智能，更容易获得欧洲用户的信任，并形成成员国之间的合力，用以构建统一的欧洲市场。最后，判定人工智能是否符合欧盟伦理，必须为其设定一套道德准则和评估标准，乃至上升至法律层面，同时也要发挥相关标准、规则与立法尝试的全球示范性力量，以此建立全球人工智能监管规则的领导权。

（二）构建软硬结合的监管框架

欧盟人工智能监管规则型领导地位的寻求还基于其强大的伦理和法律监管框架。在操作层面，欧盟将人工智能的监管规则框架分为两部分：伦理指南，即“软法”；^[1]强制性的法律法规，即“硬法”。

“软法”作为正式立法的前提，能够为人工智能的事前规制设定方向指引，目前已经成为全球人工智能监管规则的关键构成部分。一些国际性指导文件，如《蒙特利尔人工智能开发社会责任宣言》《人工智能伦理问题建议书》《布莱奇利宣言》都格外强调人工智能开发和部署过程中的“软法”力量。

欧盟在人工智能系统的研发、应用及维护中实施“软法”三步走策略：一是明确道德伦理原则的基本构成与具体执行建议；二是强调道德伦理对人工智能技术发展的功能与意义；三是以道德伦理为基点，提出立法框架。早在 2018 年，由欧盟指导，政府、学界、工商等多界代表共同成立的人工智能

[1] 称伦理指南为“软法”，因其并无法律约束力，亦无强制执行力，但其具有规范作用，对监管机构或有影响力、信誉良好的行业团体发布的原则具有一定的激励作用。“软法”具有选择适用性，其制定出台无需经过繁杂冗长的立法程序，后续修订也较为灵活。

研究机构“人工智能为人”（AI4People）就发布了《良好人工智能社会道德框架》，提出非恶意、正义及可解释性等五条道德原则，并制定了 20 条具体执行建议，^[1] 成为《人工智能法案》的基础。《人工智能法案》的提出是为了实现安全、可信和合乎道德的人工智能技术，以人为本、尊重基本权利既是该立法框架的伦理基石，也贯穿于开发和使用人工智能系统或基础模型的原则制定、人工智能系统风险划分与监督的全过程。通过以上策略，欧盟强化了“软法”的能动效应，确保道德伦理原则的基础性地位与紧密嵌入，也提升了道德伦理的监管力，将其由一般原则上升为标准乃至规则。

相较于“软法”，“硬法”更多侧重于为人工智能技术“立规划界”，予以强制性的监管。围绕人工智能立法、严格的监管要求以及统一性的监管策略，欧盟也在同步推进“硬法”实践。依托《人工智能法案》，欧盟不断推进严格的预防性立法尝试。相较于最初的草案，如今的《人工智能法案》虽然是各方妥协折衷的结果，但仍然具有严格的强制力。譬如，在由人工智能风险决定的浮动监管范围中，高风险人工智能系统面临着严格的横向要求：要符合强制性基本权利影响的评估规则、一致性评估、数据治理要求，遵循合格评定程序；要在欧盟数据库中注册，确保透明度、准确性、安全性和稳健性。满足这些要求的人工智能系统才被许可投放至欧盟市场，这些系统的提供者和用户也承担了明确且匹配的义务，以确保整个人工智能系统生命周期中对基本权利的保护和对现有立法的落实。

《人工智能法案》还设有严厉的惩罚性条款，例如，对于不遵守人工智能系统相关禁令的公司，将被处以最高 3500 万欧元或者该公司上一财年全球总营收额 7% 的罚款。^[2] 在管辖范围层面，欧盟推行统一性的“硬法”监管策略，

[1] Luciano Floridi et al., “AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations,” *Minds & Machines*, Vol.28, 2018, pp.689-707.

[2] “Regulation(EU) 2024/1689 of the European Parliament and of the Council,” Official Journal of the European Union, July 12, 2024, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689>.

人工智能系统的各方主体，包括提供者、部署者、进口商以及分销商等，都有相关义务与责任。除此之外，欧盟将法案定位为全领域通用性立法，除服务于科学研究的人工智能系统之外，各类人工智能系统都要受其统一管辖。^[1]

欧盟的“软法”与“硬法”在法律效力、内容侧重、政策实施上呈现互补关系，正如欧盟在《可信人工智能伦理指南》中规定的那样，人工智能产品满足欧盟的“软法”并不排斥其遵守法律法规的义务，且欧盟“软法”并非排他的评估标准，可与现有的行业准则互鉴与融合，在为人工智能的供应商及立法者提供参考、引领行业健康发展的同时，又不至于像“硬法”一样过度抑制创新。^[2]

（三）构建级别分明的风险体系

首先，在人工智能监管规则体系的构建中，确立技术风险分级成为欧盟的核心任务。欧盟在监管过程中，以风险为出发点，将人工智能系统分为“不可接受的风险、高风险、有限风险、极低风险”四类，针对每一类人工智能系统采取不同的监管策略，并配以具体的制度使得监管可执行落地。具体而言，针对不可接受级风险的人工智能系统，欧盟禁止任何企业或个人对其进行部署。针对高风险人工智能系统，欧盟允许相关主体在履行事前评估等义务后将其投放市场或投入使用，并进行事中、事后的持续监测。针对有限风险的人工智能系统，欧盟虽然不要求相应主体履行事前评估等义务，但其仍需遵循相应的透明度义务。针对极低风险的人工智能系统，相应主体则可自由部署和使用。^[3] 这种“基于风险等级”的灵活分类方式会随着人工智能技术的发展不断更新和补充。

其次，为了推进《人工智能法案》的实施，欧盟还采取了相应的配套措施。

[1] 李胜、高超星：《欧盟 AI 法案立法观察》，2023 年 5 月 14 日，<https://www.hankunlaw.com/upload/portal/20230518/f3ce659dbd57842f57380510bb6712dd.pdf>。

[2] 吕蕴谋：《欧盟人工智能治理的规范》，《国际研究参考》2021 年第 12 期，第 17 页。

[3] 郑孜青：《解读欧盟〈人工智能法案（审议稿）〉》，《中国外汇》2023 年第 18 期，第 30 页。

从机构设置看，欧盟将设立一个新的人工智能办公室以执行和监督通用人工智能模型的新规则。^[1]从法律适用效果看，欧盟模仿美国司法管辖中执行“效果标准”的做法扩大了管辖权范围，^[2]规定不管人工智能产品的提供者在何处，只要其生产的人工智能系统运用于欧盟区域，就需要遵守法案的要求。这与《通用数据保护条例》的法律适用原则是一致的，即最大程度上保护欧盟公民的权益。由此能够看出，欧盟意欲仿效《通用数据保护条例》，通过明确人工智能伦理、法律等治理方式提升其在全球人工智能治理的主导权，成为人工智能监管规则的全球领跑者。

最后，欧盟也在积极探索风险防范与技术创新之间的平衡之道。例如，欧盟提出支持人工智能创新的措施，即建立沙盒机制，该机制允许人工智能企业在政府部门的指导和监督下，为发展试验性和创新性的人工智能技术提供项目试验环境，在低风险、受监控的环境中进行开发测试。而且，欧盟将建立监管沙盒的权力下放到成员国层面，给予各成员国政府更多的自由裁量权。^[3]

总体而言，欧盟从伦理基础、监管原则、风险体系等三个方面入手，期望为全球人工智能监管树立“欧盟模板”与“欧盟标杆”，尤其是《人工智能法案》正成为欧盟寻求数字领导地位过程中着力打造的重要品牌。

四、欧盟政策推进的影响与限度

欧盟追求人工智能监管规则型领导地位的目标，是理想与现实妥协的产物。一方面，欧盟的规则监管设计源自欧洲长期的“道德高尚感”和“人权

[1] “EU Artificial Intelligence Act: Nearing the Finish Line,” December 15, 2023, <https://www.insideprivacy.com/artificial-intelligence/eu-artificial-intelligence-act-nearing-the-finish-line/>.

[2] 戚凯：《美国“长臂管辖”与全球经济治理》，《东北亚论坛》2022年第4期，第68页。

[3] 中国兵器工业集团第210研究所：《欧盟发布〈人工智能法案〉第一版修正案》，《机器人技术与应用》2022年第4期，第1页。

至上观”，强调在“以人为本”的基础上追求人工智能的规范化、人性化发展，展现了欧盟对技术创新与社会伦理平衡的关切，体现了其在人工智能发展中的道义感与责任感。但是，另一方面，欧盟又是极为现实主义的，它追求领导者地位的目标与措施都充满了投机性。由于技术层面的劣势，欧盟希望“失之东隅，收之桑榆”，将引领人工智能监管规则视为参与此轮数字竞争的关键入场券与重要筹码。当前，全球人工智能的新赛场正在逐渐形成，欧盟的观念与实践正在产生重要的本土与国际影响。

诚然，欧盟对人工智能监管的积极推进提升了欧盟在相关领域的影响力。欧盟创设风险分类标准和系统性监管规则体系，为业界提供了一种值得参考的监管方案，树立起了人工智能监管先行者的示范形象。欧盟可放大“布鲁塞尔效应”，积极推广自己的理念与方案，构建以欧盟规范为基准的“朋友圈”，进一步引领全球人工智能监管规则与标准的设定。不过，欧盟的政策也有着多重负面效应。第一，从欧盟内部来看，其增加了在欧企业的合规运维成本。根据数据创新中心 2021 年的测算，在欧盟管辖区内部署高风险级人工智能系统的中小型企业将面临高达 40 万欧元的合规成本，到 2025 年，合规负担将使欧盟企业每年损失 109 亿欧元。^[1] 第二，从欧盟外部来看，非欧盟的人工智能企业及其产品同样会受到欧盟监管规则的影响，它们进入欧盟市场的门槛被大大提高。^[2] 第三，从全球层面来看，欧盟的举措会刺激国际社会同侪竞相效仿，不同司法管辖区之间的监管规则壁垒可能会越来越高。

尽管欧盟对于人工智能监管规则型领导地位的追求雄心勃勃，面对国际大格局束缚、各方竞争加剧、人工智能治理结构性弊端难解、欧盟政策内生

[1] Benjamin Mueller, “AI Act Would Cost the EU Economy €31 Billion Over 5 Years, and Reduce AI Investments by Almost 20 Percent, New Report Finds,” July 26, 2021, <https://datainnovation.org/2021/07/ai-act-would-cost-the-eu-economy-e31-billion-over-5-years-and-reduce-ai-investments-by-almost-20-percent-new-report-finds/>.

[2] “Experts React: The EU Made a Deal on AI Rules. But Can Regulators Move at the Speed of Tech?,” December 11, 2023, <https://www.atlanticcouncil.org/blogs/new-atlanticist/experts-react/experts-react-the-eu-made-a-deal-on-ai-rules-but-can-regulators-move-at-the-speed-of-tech/>.

缺陷等痛点，欧盟政策的推进与落实仍存在诸多不确定性。

首先，欧盟对人工智能监管规则型领导地位的追求，无法摆脱国际竞争态势恶化的时代大局禁锢。当前，世界处于新的动荡变革期，全球安全形势愈加复杂紧张，地缘政治争斗、民粹主义、保护主义不断回潮，阵营对抗与零和博弈逆流涌动。此轮人工智能的跃迁，非但没有缓解上述难题，反而加剧了科学技术政治化、民族化、安全化的挑战。欧盟追求人工智能监管的规则型领导地位，本质上是要为自己寻求“一极”权势，因而必然会成为当下全球权力竞斗的一员。

其次，各方竞争加剧，欧盟方案“一枝独秀”或领先的时局背景正在迅速消失。不同国家或国家联盟的人工智能所处阶段、发展特点与未来规划差异极大，为应对人工智能衍生风险与挑战，中国、英国、新加坡等都在推进各自的议程。对欧盟而言，现行的人工智能监管已经基本结束了“人无我有”的阶段，全球人工智能监管规则创设及机制化合作呈现多方并举、多头竞争态势，中国与“全球南方”便是其中的关键力量。可以看到，中国坚持以人为本、智能向善，在联合国框架内加强人工智能规则治理，^[1]不断完善分类分级监管的精准治理体系，为人工智能发展方向提供了建设性思路；“全球南方”国家希望弥合数智鸿沟、共同应对人工智能风险挑战，中国先后提出的《全球人工智能治理倡议》《人工智能能力建设普惠计划》已经成为“全球南方”国家合作解决人工智能衍生问题、改善治理赤字的重要蓝本。

另外，由于一些美西方国家抱有零和式的非良性竞争思维，欧盟也很难保持“人有我优”的先发优势。技术发展与地缘政治经济竞争不可分割，这种关联增强了人工智能监管规则之争的零和博弈色彩。美国、英国等人工智能科技强国兼地缘政治大国绝不会容忍欧盟规则的“一枝独秀”，它们将会利用自身优势，分化或遏制欧盟方案在人工智能监管领域的影响力。

[1] 《习近平向联合国贸易和发展会议成立 60 周年庆祝活动开幕式发表视频致辞》，中国政府网，2024 年 6 月 12 日，https://www.gov.cn/yaowen/liebiao/202406/content_6956945.htm。

再次，欧盟方案未能解决全球人工智能治理的结构性难题。全球人工智能治理存在着三重难题：一是多利益攸关方的主体难题；二是风险和安全监管与技术创新和发展间的平衡难题；三是治理模式与机制的碎片化难题。欧盟虽然对其监管方案大力推广，但实际上也未能突破这三重难题的制约。在主体难题层面，欧盟未能做到不同主体间的利益兼顾，导致多利益攸关方呈现不同反应。譬如欧盟公众高度关注自身隐私与权利保护议题，因此对人工智能监管持广泛支持态度，^[1]但欧盟境内的大型企业却以遏制创新、阻碍外来投资、决策草率为由反对人工智能监管，西门子、空中客车等本土科技巨头曾联合发布公开信表达对《人工智能法案》的“严重担忧”，认为其会危及欧洲的竞争力。^[2]在平衡难题层面，虽然《人工智能法案》最终落地，但监管与发展的争议已经深植于立法过程中，后续仍会不断干扰相关政策落地。在碎片化难题层面，欧盟《人工智能法案》设立的高标准反而使英、美等国进一步强化了自成一体的决心，客观上加剧了全球人工智能治理碎片化的困境。^[3]

最后，欧盟现行人工智能监管政策存在明显缺陷。一方面，过度重视伦理基础，使得规则监管难以真正立足。在伦理生效过程中，容易出现“伦理洗白”（ethics washing）、伦理认知差异、伦理受制于利益等困境，限制伦理在人工智能监管中的效能。再加上欧盟的技术劣势，过于依仗伦理的监管难以在全球数字竞争中形成真正的吸引力。在《人工智能法案》首次被提出时，就有反对者认为重视伦理监管是一种过于注重控制而非能力建设的错

[1] Pascale Davies, “The New EU AI Act Is under Threat from Lobbyists, Experts and the Public Agree,” November 29, 2023, <https://www.euronews.com/next/2023/11/29/europes-ai-act-under-threat-by-lobbyists-experts-and-the-public-say>.

[2] Michelle Toh, “‘Serious Concerns’: Top Companies Raise Alarm over Europe’s Proposed AI Law,” June 30, 2023, <https://edition.cnn.com/2023/06/30/tech/eu-companies-risks-ai-law-intl-hnk/index.html>.

[3] Shana Lynch, “Analyzing the European Union AI Act: What Works, What Needs Improvement,” July 21, 2023, <https://hai.stanford.edu/news/analyzing-european-union-ai-act-what-works-what-needs-improvement>.

误做法，是典型的本末倒置。^[1] 另一方面，具体监管细节存在缺陷。以《人工智能法案》为例，其存在多处细节缺陷：一是缺乏细化的责任体系，对规制对象的责任界定不清晰，也未明确多方参与场景下多个责任主体间的责任分配，这可能导致责任确定困难。^[2] 二是法案在责任承担方式上较为单一，处罚方式主要是罚款，新旧立法之间的术语差异可能会诱发合规责任矛盾。例如，《人工智能法案》中使用的“提供者、部署商、进口商和分销商”等术语与《通用数据保护条例》中通常理解的数据控制者和处理者不同，这会导致两部法案在同时适用的情况下合规责任不清晰。^[3] 三是风险分级方法易僵化，人工智能跃迁速度极快，行业变化日新月异，欧盟的四级分类法无法精确捕捉人工智能风险的奇点突变特征，容易导向“不监管”或“重监管”的简化选择。四是风险分级划定错误的代价较高，如果更低等级风险的人工智能系统被判入更高等级风险范围，将面临繁重且不成比例的监管要求，系统开发者难以创新，甚至可能为避免麻烦选择直接回避欧盟市场，这不仅会剥夺欧洲民众享受人工智能科技益处的权利，还会成为科技治理破坏科技创新的反面典型。^[4]

五、结语

全球人工智能治理目前仍处于艰难的起步阶段。全球主要行为体都在积极行动，试图凭借各自优势获得人工智能领域的领导地位。欧盟进行了领先

[1] Sebastian Vos et al., “2023 Spotlight Series on Global AI Policy,” October 4, 2023, <https://www.globalpolicywatch.com/2023/10/spotlight-series-on-global-ai-policy-part-i-european-union/>.

[2] 曾雄等：《欧盟人工智能的规制路径及其对我国的启示——以〈人工智能法案〉为分析对象》，《电子政务》2022年第9期，第67页。

[3] 《全球人工智能治理大变局之欧盟人工智能治理监管框架评述及启示》，金杜律师事务所网站，2023年7月22日，<https://www.kwm.com/cn/zh/insights/latest-thinking/eu-artificial-intelligence-governance-and-regulatory-framework.html>。

[4] William Crumpler, “Europe’s Strategy for AI Regulation,” February 21, 2020, <https://www.csis.org/blogs/strategic-technologies-blog/europes-strategy-ai-regulation>.

性的规则创建，形成了以伦理为基础的监管框架与风险划分体系，然而，国际格局变动、多元监管方案之争以及欧盟自身积弊制约着欧盟方案的效力。这也恰好说明，想要实现有效的全球人工智能治理，仅靠“一枝独秀”是不可能的。

人工智能技术的政治化、泛安全化不断恶化治理环境，不同国家的治理初衷也愈加复杂，有的是为了安全而监管，有的为了竞争而监管，有的是二者兼具。不同的监管初衷塑造了差异化的行事方式。一些国家坚持智能向善发展、筑牢数智安全底线，一些国家则以人工智能引发风险挑战为借口，不择手段打压他国，隐匿性地塑造人工智能的政治光谱与政治立场，将其视为争夺数字霸权的新手段。全球人工智能治理的困局是技术革命与全球秩序重构叠加的缩影，在技术权力与地缘政治交织的漩涡中，作为纯粹生产力工具的人工智能正被异化为新型战略博弈筹码。想要破解人工智能治理僵局，亟需强化“以人为本”的伦理共识，只有让技术进步回归增进人类福祉的本源，才有可能切实推进全球人工智能协同治理。在此基础上，提升技术互信，建立更具包容性的治理框架，或能将当前各自为政的“数字堡垒”转化为互联互通的“智慧网络”。

【责任编辑：肖子璇】