全国中文核心期刊
中文社会科学引文索引（CSSCI）来源期刊
中国人文社会科学期刊AMI综合评价（A刊）核心期刊
中文科技期刊数据库（全文版）收录期刊
RCCSE中国核心学术期刊（A）

2025. 3
VOL.18 NO.3

公共行政评论

JPA
Journal of Public Administration

公共行政评论
Journal of Public Administration

全国中文核心期刊
中文社会科学引文索引(CSSCI)来源期刊
中国人文社会科学期刊 AMI 综合评价
(A 刊)核心期刊
RCCSE 中国核心学术期刊(A)
中文科技期刊数据库(全文版)收录期刊
中国核心期刊(遴选)数据库收录期刊
中国学术期刊(光盘版)收录期刊

6 月 15 日出版　　2025 年第 3 期　总第 105 期

●专栏：政策执行研究

确定性与不确定性：公共政策的相机执行策略选择研究
…… 李胜会　李　丹　朱绍棠　1

地方社会政策执行的内在逻辑
——以“多校划片”义务教育改革为例
…… 曾　渝　蒋光明　20

●论文

专项整治治理基层数字形式主义何以有效？
——基于整治主体的分析
…… 张靖宇　甘　甜　过　勇　37

理解城市治理敏捷转型的复杂性及其运作逻辑

——基于解构制度性行动情境的研究视角

…… 陈廷佳　赵紫懿　孙　涛　57

基层治理中历史遗留问题何以产生？对政府决策中“当下主义”策略的案例分析

…… 余成龙　郑晓华　78

激活“公”与“共”：占据结构洞的县域社会组织

…… 林雪霏　单雅杰　98

信息干预与新业态从业人员养老保险的参保意愿：基于外卖骑手随机实验的

实证研究　…… 申梦晗　苏楚平　申广军　119

城市政府合作关系的不均衡性：基于长三角官员异地考察网络的研究

…… 林钢健　徐元朔　张蔚文　138

全球人工智能治理中的联盟冲突与政策学习

——基于倡议联盟分析框架　…… 雷晓康　于林霞　157

●理论综述

公共管理研究中的元分析方法：进展、范式与方向

…… 苗　青　朱星魁　175

●英文目录与摘要 …… 196

全球人工智能治理中的联盟冲突与政策学习

——基于倡议联盟分析框架

雷晓康　于林霞*

【摘要】全球人工智能治理已超越单纯的技术规制范畴，演变为涉及人类文明伦理基底与国际政治经济秩序重构的深层博弈。作为具有显著非中立性特征的战略技术，人工智能（AI）正成为大国价值体系竞争的新型载体，其治理主导权争夺已成为主要经济体的战略优先事项。然而，技术迭代的不确定性叠加地缘政治竞争，导致国际治理生态呈现“规则碎片化”与“价值阵营化”的双重异化。“强监管联盟”与“弱监管联盟”在技术伦理、规则制定权与跨境管辖等维度存在根本分歧，现有治理机制难以化解这些结构性矛盾引发的协调困境。然而，“智能向善”伦理范式、多层治理实践及国际组织构建的专业对话平台，为突破治理僵局提供了可能。基于倡议联盟框架，研究建议构建三级协同机制：核心层强化制度性履约、辅助层优化动态协调、影响层完善参与式传播。这种立体化设计既能确保技术发展的人类中心导向，又可增强治理方案的地域适应性，最终达成科技伦理与人类价值的正向循环。

【关键词】人工智能治理　倡议联盟框架　强监管联盟　弱监管联盟　政策学习

【中图分类号】D63　　【文献标识码】A

【文章编号】1674-2486（2025）03-0157-18

一、问题的提出与政策演进

（一）全球人工智能治理问题的缘起

人工智能（AI）与人类之间相互促进的关系，不仅提升了我们处理任务和

* 雷晓康，西北大学公共管理学院教授。于林霞，西北大学公共管理学院博士研究生，渭南师范学院经济与管理学院讲师。

基金项目：陕西省哲学社会科学重大理论与现实问题2021年度重点智库研究项目“数字赋能政府公共服务高质量发展及实现路径研究”（2021ZD1048），四川省哲学社会科学重点实验室重点项目“人工智能等先进技术在重大突发事件应急处置中的应用实践”，陕西省教育厅2020年一般专项“创新驱动战略下我国西部地区先进制造业发展路径研究”（20JK0121）。

解决问题的能力，还放大了我们的智慧、创造力以及道德判断力。根据《2024人工智能安全报告》，AI技术的恶意使用急剧增加，2023年基于AI的深度伪造欺诈增长了3000%，基于AI的钓鱼邮件增加了1000%。此外，多个国家背景的APT（Advanced Persistent Threat）组织利用AI实施了超过10起网络攻击事件，在政治、网络、物理和军事安全方面构成了重大威胁①。美洲地区的种族歧视问题因技术的偏见和误用变得更加尖锐，南北世界之间的贫富差距由于区块链、大模型、深度学习等高科技领域的技术壁垒愈发难以逾越，智能技术“赋能”的国家安全风险、伦理道德挑战、贸易与产业竞争等传统社会问题呈指数级增长趋势，给全球治理带来了前所未有的挑战。尤其是DeepSeek等先进大模型的出现，推动了大模型训练方式的革新，显著提升了生成效率与结果的质量。然而，人类认知的局限性使得我们难以完全理解和掌控AI的决策逻辑，甚至可能面临被技术反噬的风险。随着人工智能逐渐成为社会发展的关键驱动力，我们正日益逼近“科林格里奇困境”，即在技术成熟之前难以预见其影响，而等到影响显现时，却已难以有效控制（阿扎尔，2024）。霍金曾警告：“如果不对人工智能加以限制，人类社会可能会迎来一个黑暗的未来（傅莹，2020：1）。”《人机对齐：如何让人工智能学习人类价值观》作者克里斯汀说：“机器学习表面上是技术问题，但越来越多地涉及人类问题，人类、社会和公众难题正在变得技术化，技术难题正在变得人性化、社会化和公众化（布莱恩，2023：10）。”

人工智能的治理是一个全球性问题，需要各参与主体进行复杂而漫长的谈判，其首要条件是共有观念的形成，且只有共有观念进入各方均认同的相关国际文本，如公约、条约、协定或协议，才能保证其影响力不至于在利益的博弈中消亡于国际社会。也因此，联合国于2016年推动国际社会谈判全球人工智能治理的问题，开始从制度规范等实践层面进行应对，拉开了全球人工智能治理的序幕。

（二）全球人工智能治理政策的诞生与博弈

全球人工智能治理的严肃性与复杂性使其参与主体从主权国家向智能企业、非营利组织以及个人发生多元化的转变，政策的约束性也经过了从松散到严格的渐进式强化过程，这一演变轨迹受到若干焦点事件的推动，相关研究成果和演进脉络如图1所示。

① 参见奇安信发布《2024人工智能安全报告》，2024年2月29日。

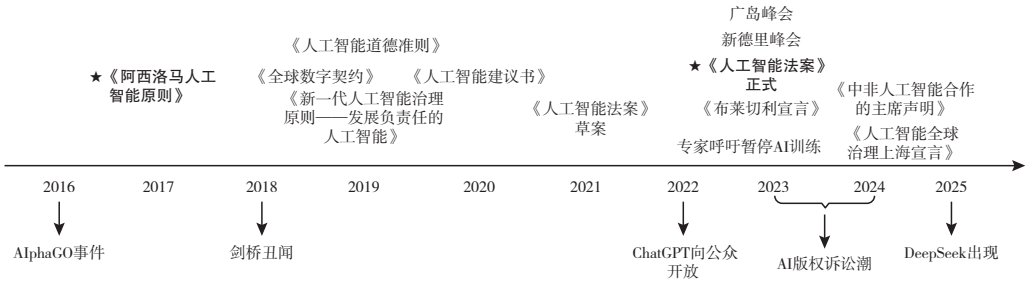

图 1　全球人工智能治理中的焦点事件与政策演进

资料来源：作者自制。

1.《阿西洛马人工智能原则》的签署（2017 年）

2016 年，AlphaGo 以 4∶1 的成绩战胜围棋世界冠军李世石，这一事件不仅展示了 AI 技术的巨大潜力，也引发了全球对 AI 能力及其伦理挑战的广泛关注。这是首次向公众直观展示 AI 的突破性进展，激发了社会关于 AI 伦理及失控风险的早期讨论，并促使科技界提出了“AI 安全”的议题，成为后续治理工作的起点。次年，“有益的人工智能会议”在美国加利福尼亚州阿西洛马召开，由非营利组织“未来生命研究所”主办。此次会议汇聚了来自世界各地的人工智能专家、伦理学家、政策制定者和企业家，包括埃隆·马斯克、斯蒂芬·霍金以及 DeepMind 的联合创始人肖恩·莱格等。会议上达成了《阿西洛马人工智能原则》，强调 AI 的发展应服务于人类共同利益，确保人类对 AI 技术的控制权，避免技术失控的风险，并提倡优先资助 AI 的安全性、伦理性和社会影响研究。这些共识为后续全球人工智能治理政策的推进奠定了坚实的基础。此后，多个治理主体纷纷出台相关政策措施，认可并遵循上述共识，例如，联合国的《全球数字契约》，欧盟的《人工智能道德准则》，中国的《新一代人工智能治理原则——发展负责任的人工智能》。此外，经济合作与发展组织（OECD）发布的《人工智能建议书》等文件也在不同程度上支持了技术向善的基本理念。特别是在 2021 年，联合国教科文组织（UNESCO）推出的《人工智能伦理问题建议书》，作为首个全球性的 AI 伦理框架，提出 AI 的发展应遵循保护人权、透明度、问责制和环境可持续性等原则，为各国在国家层面制定 AI 伦理法规提供了重要参考。

2.《人工智能法案》的出台（2023 年）

智能技术的迅猛发展，为教育、医疗、交通等领域带来了巨大的变革和赋能，但同时也暴露了许多问题。例如，2018 年的剑桥分析丑闻揭示了数据滥用的风险，2022 年 AI 算法对多国选举的干预凸显了算法操纵的危害，2023 年至 2024 年间生成式 AI 版权诉讼潮则进一步展示了信息战中的不确定性风险。这些

问题表明，仅依靠道德伦理难以有效约束人工智能技术应用中的违规行为，更加严格的规章制度亟须出台。在这样的背景下，2023 年联合国成立了由 39 名成员组成的 AI 高级别咨询机构，人工智能治理正式成为联合国全球议程的一部分。同时，国际社会加速推进制定具有强制性的全球人工智能治理规则。2023 年的广岛峰会与新德里峰会就开发先进 AI 系统应遵循的国际标准进行了深入讨论。同年 12 月，欧盟《人工智能法案》的生效，作为首部全面监管 AI 的法律，为国际社会树立了新的标杆。随后，中国、美国、英国、德国等国家共同签署了《布莱切利宣言》，强调需共同应对如大模型滥用、失控风险等全球性挑战，并同意建立“全球 AI 安全研究网络”，以促进国际合作和技术风险的研究。这是首次将“AI 系统失控风险”定义为全球共同面对的挑战，并确立了“国家主导+企业协同”的双轨制治理模式。

二、文献回顾与理论框架构建

在关于全球人工智能治理讨论中，尽管表面关注点聚焦于和平与发展，但实质触及了深层次的经济利益分配及伦理价值取向。形式主义与算法“暴政”之间的冲突、静态规则与动态技术的脱节、“制度理性与技术野性”的矛盾交织，这些都表明同步发展适宜的治理工具变得尤为紧迫和必要。部分学者认为人工智能的治理需要强有力的监管措施，他们主张采用更具刚性的制度规则，这有助于明确责任边界，减少社会风险，并利用统一标准降低合作成本，防止不良竞争（沈伟，2022；黄锫，2025）。然而，另一些观点认为，僵化的监管规则难以跟上技术快速迭代的步伐，在应对新型社会问题时往往缺乏灵活性，应当完善柔性监管机制①。还有学者建议采取“软硬兼施”的策略，结合两者优势形成混合治理模式。“中心–外围”治理模式就是一个典型例子，它以“硬法”为核心，依靠国家共同制定的硬法规制关键领域的问题，确保底线约束的同时提高 AI 系统的可靠性；而针对特定应用场景中的个性化问题，则依赖非国家行为体制定的“软法”，让接近技术前沿的专家参与制定更具灵活性的规范框架。当前人工智能治理呈现出“软法繁荣而硬法滞后”的失衡局面，这种制度错配带来了双重困境，一方面，部分名义上的“软法”在实践中异化为具有实质约束力的监管工具，其隐性制裁效应严重抑制了技术创新活力；另一方面，软法体系的弹性空间被部分主体策略性利用，衍生出新型治理的套利行为，加

① 参见 Brookings，*How soft law is used in AIgovernance*，2021 年 5 月 27 日，http://www.brookings.edu/techstream/howsoft-law-is-used-in-ai-governance/。

剧了国际的纠纷。例如，在数字产品跨国贸易中，美国曾利用 WTO 的“安全例外”条款打压华为，这实质上是变相的贸易保护主义，阻碍了跨国数字贸易的发展（朱明婷，2023）。当前全球人工智能治理已形成基础性共识，但受各国制度环境、文化传统和发展诉求等差异的影响，各治理主体在监管路径与价值取向上仍存在显著分歧。本研究将运用倡议联盟框架，系统解析不同利益联盟的政策信念体系与互动机制，以期为构建包容、均衡的全球人工智能治理体系提供理论参照。

倡议联盟框架（Advocacy Coalition Framework，ACF）是一种以信念系统为核心要素的政策分析工具，自保罗·萨巴蒂尔及其同事在 20 世纪 80 年代首次提出这一理论框架以来，ACF 因其能够细致捕捉政策制定过程中多元利益博弈的本质而受到了学术界的广泛关注与认可（保罗·A. 萨巴蒂尔，2011）。本文使用倡议联盟框架来分析全球人工智能治理问题，原因在于：其一，政策子系统的多元化与碎片化。ACF 假设政策制定发生在由多元主体（政府、企业、非营利组织、专家、国际组织等）构成的“政策子系统”中。全球人工智能治理的显著特点正是多方利益相关者的高度参与（如国家政府、科技巨头、伦理组织、国际标准化机构等），且缺乏统一的权威机构，ACF 能够有效刻画这种碎片化结构中的联盟互动。其二，政策信念系统的冲突。ACF 强调政策争议的本质是不同联盟围绕“深层核心信念”（如价值观、伦理立场）和“政策核心信念”（如具体政策目标）的冲突。全球人工智能治理的核心争议（如 AI 发展的“安全”与“创新”以及“经济利益”与“人权保护”等）正是这种信念冲突的典型体现。其三，政策学习与外部事件驱动政策变迁。ACF 认为政策变迁缘于联盟间的政策学习或外部冲击（如技术突破、重大事件等）。AI 领域的技术迭代速度极快，且关键事件频繁冲击既有治理框架，这与 ACF 的动态分析逻辑高度吻合。本研究旨在运用 ACF 理论探讨以下问题：在全球人工智能治理领域中，形成了哪些有影响力的联盟组织？这些联盟之间存在怎样的冲突与分歧？它们如何通过相互间的政策学习来突破政策僵局？

三、全球人工智能治理中的联盟冲突

“AI 教母”李飞飞（2024：10）曾发出警告：“人工智能正在成为一种特权，一种排他性极强的特权。”在这种涵盖全球范围的智能技术生态系统中，没有任何一个国家或机构能够独立解决其中涉及的所有问题，相应地，“行为准则”的制定同样需要各国政府、国际组织以及非政府组织之间的通力合作。具体而言，随着各主权国家开始参与人工智能的发展与应用，并成为这一领域的

重要参与者，现实世界中的国际合作与竞争态势也开始逐步渗透进人工智能领域，这要求新的治理框架在构建全球人工智能秩序方面必须发挥关键作用。

（一）人工智能治理问题的严肃性使联盟冲突凸显

人工智能技术并未超越技术的非中立性，而是成为大国或国家联盟价值理念之争的新兴载体。AI作为工具本无倾向，但“以人为本”的治理宗旨呼吁人工智能治理框架需确保人工智能技术发展始终与人类发展目标深度契合，这就为AI注入了“人本属性”（陆卫明，2025）。技术不再仅仅是工具，更是塑造国家、企业乃至社会发展路径的关键变量。在人工智能时代，技术权力进一步分解为工具性权力、话语性权力和制度性权力。三种权力维度共同刻画了人工智能时代技术权力的复杂内涵，并通过多重路径对国家安全与全球治理产生深远影响。在人工智能技术的跨国部署中，技术控制者巧妙地将本国的文化价值与意识形态融入全球知识体系，并借助所谓“透明性”与“公平性”等治理理念推广其技术模型，导致这些技术的海外受众在无形中接受并内化来自外部的叙事话语。就技术标准层面而言，人工智能的算法设计原则、数据安全规范和技术评估框架逐步成为国际治理中的核心议题。这一过程往往由技术实力雄厚的国家与行业领先企业所引领。技术标准一旦被确立并广泛采纳，将对全球技术应用的格局与发展轨迹产生深远而持久的影响。例如，在自动驾驶、社交媒体和数据隐私领域的技术规范方面，领先企业和先发国家可通过技术标准的制度化巩固对技术市场的控制能力。而在伦理软法的构建层面，相关行为体能够利用技术与影响力上的优势，将其立场与偏好转化为对治理伦理框架的影响力，并将自身利益深度融入这些伦理准则之中。然而，技术规则与伦理软法的制定权高度集中在少数国家和领先企业，从而在实质上构成了制度性权力的不对称，这也导致相对弱势国家不得不被动地接受技术主导者设定的治理框架（孙志伟，2025）。

（二）基于信念系统差异的跨联盟冲突

随着人工智能技术的迭代加速，全球治理主导权已成为主要经济体的战略竞争焦点，各主权国家围绕技术主权与利益诉求，相继推出差异化治理方案，但技术风险的不确定性与地缘政治博弈相互交织，导致国际治理格局呈现“规则碎片化”与“阵营化”特征。在此背景下，逐渐形成以规制强度为分野的两大联盟体系（见表1），“强监管联盟”主张通过具有法律约束力的刚性治理工具建立统一的规制框架，强调通过制度化约束防范系统性风险；“弱监管联盟”推崇非强制性的治理工具，依靠市场自律与技术自治实现弹性治理，主张通过沙盒机制平衡安全与发展。

表 1　全球人工智能治理主要的联盟体系

联盟类型	强监管联盟	弱监管联盟
代表团体	中国、欧盟	美国、印度、部分亚太国家
民族理念	人权至上	技术至上
价值理念	伦理优先	创新优先
价值取向	社会取向	政治取向+经济取向
主要价值诉求	在确保人类安全的前提下发展人工智能技术	充分发展智能技术，获取更多的经济利益
利益诉求	谋求安全与和平的发展	成为掌控智能技术的国家，谋求国际政治地位，掌握话语权
针对生成式 AI	要求对深度伪造、版权侵权等情况实施事前备案和履行透明度义务	主张通过事后追责和行业标准代替政府审批
数据与算力相匹配	要求建立全球数据共享机制和算力补贴	强调知识产权保护，反对强制技术转移
全球规则制定权	多边主义联盟，倡导以联合国为中心的统一规则	倾向区域或国家主导的碎片化规则

资料来源：作者根据沈伟（2022）、毕超（2024）、陆卫明（2025）等整理。

1．人工智能技术研发的信念差异

在人工智能全球治理的实践探索中，表层技术争议的本质是深层次价值理念的碰撞。“弱监管联盟”秉持市场原教旨主义的治理哲学，其核心主张可解构为三个维度。其一，在治理范式层面奉行“技术自治”原则，主张通过市场机制的自发调节实现资源配置，采用“负面清单”管理模式，仅在重大公共利益受损时实施有限干预，美国国家标准与技术研究院（NIST）发布的《人工智能风险管理框架》就是采用这一自愿性的合规机制。其二，在价值体系层面，以工具理性为主导思维，将 AI 技术视为维持霸权地位的“战略杠杆”，美国通过的《芯片与科学法案》旨在构建技术封锁联盟。其三，在战略文化层面体现“技术决定论”思维，信奉“技术代差=战略优势”的零和博弈逻辑。美国前司法部部长威廉·巴尔（William Barr）在一场演说中称：“自 19 世纪 90 年代以来，美国一直是全球技术领袖。美国的机会和现有的繁荣及安全均建立在拥有领先技术的基础之上。”① 回顾人类历史长河，技术进步一直是战争形态变化、武装力量升级、军事战略改良的主要推动力量，也是国家硬实力的重要基础和

① 参见 https://www.justice.gov/opa/speech/transcript-attorney-general-barr-s-remarks-china-policy-gerald-r-ford-presidential-museum。

支撑。尚武精神强烈的美国更将谋求和保持技术的绝对优势奉为圭臬，“技术至上”的民族理念主导其在各领域寻求技术领先和统治地位（徐菲，2024）。美国政府也积极推动“国防 AI 能力建设和治理”，体现其希望通过人工智能增强国防实力和全球领导力的“刚性”意图。

弱监管联盟虽标榜柔性治理，实则暗含监管霸权逻辑。美国的人工智能治理政策呈现出鲜明的“战略利己主义”特征，其规则体系的构建始终服务于维护国家竞争优势与巩固技术霸权。2022 年，美国专门出台《芯片与科学法案》，以便政策直接干预芯片产业（王威力，2025）。同年，美国国防部发布《负责任的人工智能战略和实施途径》，表面为开发智能的、弹性的、可靠的下一代网络技术提供伦理指引，实则为实施“马赛克战”计划打掩护，以便利用智能技术打造强大的战场全息数字化能力。2023 年 10 月，美国发布《关于安全、可靠和可信的人工智能的行政命令》，强调人工智能技术需以美国公民和国家利益为先。2024 年 3 月，美国众议院通过《保护美国人免受外国对手控制应用程序侵害法》法案，迫使字节跳动剥离 TikTok 应用，这暴露了美国在高科技和人工智能领域的霸权心理。除此之外，美国的“国家人工智能系统的安全与风险管理”集中于生物识别、深度伪造等技术领域，重点规避技术在安全与伦理上的潜在威胁，反映其通过法律框架与伦理规范引导技术发展的倾向（Yudkowsky，2008）。

强监管联盟秉持多边主义理念，更期待各方携手，尽力纾困人工智能跃迁带来的数字鸿沟扩大的难题，同时以全人类智慧防范人工智能“叛变反噬”等根本性风险。比如，欧盟强调战略自主，希望通过“欧洲地平线计划”、支持量子计算以及边缘计算等大规模投资，驱动人工智能技术进步，减少对外部技术的依赖。同时，欧盟积极构建完善的伦理和法律体系，以及个人数据保护体系，彰显其在世界 AI 治理领域的影响力与领导力（王彦雨，2024）。中国旨在通过技术创新深度重构社会服务体系。在公共治理层面，依托人工智能优化资源配置效率，通过政务大模型提升基层治理精准度；在产业转型方面，利用自动驾驶技术实现智能规划物流路径，降低运输成本；在可持续发展领域，借助 AI 算法大幅提升气候预测精度，为环境治理提供科学支撑。通过技术与应用场景的深度融合，我国实现了社会服务智能化水平的系统性提升。

2．国际话语权的争夺

全球人工智能治理场域中，弱监管联盟正通过制度性权力的策略性延伸，实施多维度联盟扩张策略，“俱乐部模式”是其惯用的手段，七国集团（G7）通过《广岛 AI 进程》建立了全球治理的西方中心主义规则体系，并在 AI 伦理准则中嵌入符合美企利益的条款。2023 年的数据显示，美国主导的标准组织贡

献了62%的全球AI技术规范。技术联盟等级化固化了国家之间的“技术代差”，处于核心圈的“五眼联盟”国家拥有完全的技术共享，次级圈的欧盟国家有条件地接受技术转移，外围圈的南方国家仅获得阉割版的技术方案，《全球跨境隐私规则》体系中的差异化准入安排佐证了这一理论。美国作为弱监管联盟的典型代表，试图将技术标准武器化、联盟关系工具化，构建具有技术殖民主义特征的全球人工智能治理秩序。一方面，美国以人工智能技术合作为名，在技术研发领域积极争取盟友支持。自2020年起，美国陆续与英联邦国家形成“共同的人工智能愿景”，与印度启动合作研发项目，同澳大利亚联合资助公平和可信赖AI的研发，以及与欧盟在多个应用领域开展AI合作（贾开，2024）。另一方面，美国又以维护和平发展为旗号，构建多元化的排他性联盟。在数字贸易治理领域，美国建立了印太经济框架（IPEF）、美国与欧盟贸易和技术委员会（TTC）、芯片四方联盟；在信息保护领域，美国与日本就加强全球跨境隐私规则（CBPR）合作、推动实施该规则和处理者隐私识别系统等事项达成共识。此外，随着中美国际关系的日益紧张，美国加强了对华竞争态势，并借助“五眼联盟”、七国集团等平台进行联合围堵。特别是2024年的“海底冷战”，美国试图攫取海底通信电缆铺设的控制权，在联大会议期间故意拉拢16国及地区共同声明要对海底电缆进行“联合管控”，以保护数据传输中的安全，并企图将中国排除在外，美国此举无疑是在全球数据传输领域构建围堵、孤立中国的联盟（王海滨，2025）。

相反，强监管联盟虽然表面呈现出刚性监管的姿态，但在实践中展现出显著的治理弹性，其政策执行往往融合了协商共治与渐进调适的柔性特质。例如，欧盟虽然在其域内法层面率先推出被称作史上最完备的《人工智能法案》，在与新加坡、日本和韩国等签订的数字经贸协定中，并未涵盖或推广其域内人工智能治理的经验和做法。《欧盟-新加坡数字伙伴关系协定》（*EU-Singapore Digital Partnership*，EUSDP）中包含的“AI-specific”规则仅限于强调提升成员方人工智能治理系统的互操作性的意义。这种“刚柔并济”的治理范式，既保持了必要的监管威慑力，又通过动态评估机制为技术创新预留了发展空间。阿努·布拉德福德（Anu H. Bradford）曾将欧盟偏好的规则输出方式称作“布鲁塞尔效应”（the Brussels Effect），即以单边监管为基础，通过市场机制推动制度外化，借助市场力量敦促其他经济体接受和应用其规则（Bradford A，2020）。中国坚持独立自主外交政策，在人工智能治理这一国际议题中的强监管态度旨在规范国际秩序。我国已加入的《区域全面经济伙伴关系协定》（RCEP）及正在推动加入的《全面与进步跨太平洋伙伴关系协定》（CPTPP）和《数字经济伙伴关系协定》（DEPA），均强调建立合理的跨境数据规则的重要性，这本质上是为主

动对接国际经贸规则做准备（张亮，2025）。

3.“长臂管辖”的冲突

弱监管联盟利用治理的规则漏洞，行“长臂管辖”之实，主要表现为“弹性条款”的巧妙利用。美国将 OECD 的《AI 原则》中关于“合理例外”的内容扩大解释为制度性豁免，在《美加墨协定》的数字章节中植入技术性壁垒条款。

利用司法权向企业施压是美国“长臂管辖”的一个主要手段。在 2013 年的微软案件中，美国司法部要求总部在美国的微软提供其在爱尔兰都柏林数据中心存储的用户电子邮件数据，但微软公司认为该数据存储的服务器在爱尔兰都柏林，美国属于“长臂管辖”，拒绝服从该命令。受迫于美国执法部门在国际社会的强势地位，微软败诉。2018 年 3 月，美国总统特朗普签署了《澄清域外合法使用数据法》，该法案的出台使得美国调取本国公司储存在境外的数据合法化。这意味着美国通过司法手段创建了一个关于数据主权划分的新标准，以此打破各国数据本地化政策的数据保护屏障，本质上是美国的世界霸权行径。

遏制中国的技术发展是美国“长臂管辖”的一个重要目的。2024 年 2 月底美国白宫发布《关于防止受关注国访问美国人大量敏感个人数据和美国政府相关数据的行政命令》（以下简称《第 14117 号行政命令》）①，成为限制美国涉华数据出境的总纲领；同年，美国全力推进“清洁网络计划”，并以互联网数据基础设施的监管为重点，打压排挤中国通信公司的发展。一方面，美国的电信服务部针对海底电缆出台了更高的网络安全标准，同时美国网络安全与基础设施安全局发布严苛的数据安全标准规定，以涉及中企及相关人员的数据访问及使用限制为重点；另一方面，美国以安全调查为由，限制在华企业等外国厂商对美国云服务（IaaS）产品的访问，并限制被美方认定有“数据风险”的 TikTok、天猫商城等软件的应用（王海滨，2025）。

（三）各国利益博弈加剧联盟内部的冲突

尽管弱监管联盟在治理理念和基本立场上保持统一，但联盟内部在涉及国家利益和经济利益等实质性议题时，往往会出现明显的立场分化与政策冲突。例如，印度就曾明确表示反对美国制定高标准的数字贸易规则，主张实现数据管理“本土化”和数据保护“独立化”（张东冬，2025）。联盟集团内部的冲突也同样激烈，七国集团（G7）成员国围绕 AI 监管分化为两大阵营——以法国、德国和意大利为代表的严格法律途径派，以及美国、日本和英国组成的灵活监管策略派。前者借鉴《人工智能法案》模式，通过广泛的法律法规来规范 AI 发

① 参见 https://www.eol.cn/news/lzmzl/202504/t20250411_2663176.shtml。

展；后者主张识别高风险领域如自动驾驶、医疗诊断等，并量身定制具体的监管方案，同时使用指导原则、行业标准等柔性工具引导企业合规经营，保持技术创新活力（贾开，2024）。尽管美国与日本是长久以来的战略联盟，但是在人工智能治理问题上也各持己见，美国倾向于采用一种分散式的治理模式，强调创新优先。这种模式依赖于市场的灵活性和企业的自我监管，政府的角色更多是提供指导和支持，而不是直接干预。日本选择了一种较为宽松的软法治理模式，旨在促进技术的快速发展。

强监管联盟虽然普遍认同硬性约束对维护国际秩序的作用，但受各国政治体制差异、文化传统分野等因素影响，其内部在治理理念架构与政策实施路径上仍存在显著的分歧。在风险的分级分类管控方面，欧盟的《人工智能法案》确立三个层次的风险系统，包括禁止使用的 AI 系统、高风险的 AI 系统以及须履行透明度义务的 AI 系统。中国的法规同样基于风险等级构建合规要求，但中国现行的法规并未正式规定“禁止”技术清单，而是对具有舆论属性和社会动员能力的服务加以限制。欧盟因其内部成员国众多，在人工智能这一公共性议题上面临众口难调的问题，关于对“基础模型”的监管，欧盟委员会、欧洲理事会和欧洲议会之间都存在分歧，欧洲议会选择对基础模型施加更严格的透明度要求，而欧洲理事会则倾向于给予更多灵活性，并建议由欧盟委员会在未来为这些模型设定具体规则。在行业友好性及创新保护方面，随着欧洲本土科技公司的发展，部分成员国担心过于严厉的规定可能会抑制本国企业的创新能力。因此，法国、德国和意大利等国倾向于支持一个对行业更为友好的框架，以促进而非限制 AI 技术的发展。

四、全球人工智能治理满足联盟政策学习的三个条件

（一）全球人工智能治理子系统内的冲突是可调解的

在技术正义理论视域下，全球人工智能治理子系统呈现冲突可调解性的理论特征。第一，价值对齐机制是治理体系的内生稳定器。技术哲学研究表明，机器智能系统必须通过价值嵌入实现与人类文明基准线的伦理校准。联合国教科文组织《人工智能伦理建议书》提出的“智能向善”原则，实质上构建了跨文明的价值公约数。这种对齐过程不仅需要技术层面的算法价值观审计，更需建立包含多元文明伦理准则的全球价值图谱。第二，透明度互构机制是信任生产的物质基础。通过构建跨国技术审计联盟（如欧盟 AI 监管沙盒计划）、实施研发信息披露分级制度（参照《瓦森纳协定》技术转移机制），以及建立风险

评估协同平台（如 OECD 人工智能政策观察站），可将技术黑箱的认知偏差大幅降低①。这种制度安排使贝叶斯博弈模型中的信息更新效率提升至帕累托最优水平。第三，弹性治理体系确保制度供给的动态适配。采用“硬法-软法-技术标准”的三层架构设计，允许治理主体根据技术成熟度和发展梯度选择规制工具组合。构建“价值对齐-透明互信-弹性规制”的复合治理范式，能将 AI 技术竞争约束在生产力发展的帕累托前沿，最终实现技术创新、文明存续与物种正义的共时性发展。这为破解人工智能时代的“科林格里奇困境”提供了制度解方。

（二）全球人工智能治理问题是可分析的

多角度构建的理论框架为全球人工智能治理奠定了学术基础，而国际合作中的治理实践则探索了实现路径。两者相辅相成，共同推进了这一公共领域问题的有效解决。

从理论框架的构建上，“价值链治理模型”强调了智能技术发展的“人本属性”，关注如何通过治理 AI 技术研发、数据流动、算法部署和产业应用等关键环节，使 AI 发展具有公平性、安全性和可持续性（何宇，2020）。“敏捷治理模型”强调快速反馈循环和动态政策调整，建议采用实验性监管，尽可能地让国家、社会、个人等多方利益相关者参与治理（杨胜刚，2025）。“风险管理金字塔模型”提出了风险分层的治理框架，个体层面的算法偏见、组织层面的企业数据滥用、国家层面的 AI 军备竞赛，三个层面的风险治理需对症下药方能从源头上化解智能技术带来的系统性风险②。“博弈论驱动的合作治理模型”认为各主权国家在 AI 发展中的竞争可能导致集体非最优结果，应该通过协调博弈建立国际标准以避免冲突，并通过设置激励相容的治理结构让更多的治理主体参与进来（Dafoe A，2018）。

在实践层面，国际社会已在隐私保护、算法公平等方面达成多项共识，如《联合国教科文组织人工智能伦理问题建议书》、欧盟的《通用数据保护条例》（GDPR）等，尤其是 2022 年欧盟提出的《人工智能法案》要求对 AI 生成内容进行标识，得到了国际社会的广泛认可。分领域的治理也取得了一定的进展，国际医疗器械监管机构论坛（IMDRF）《关于软件作为医疗器械的风险分类指南》（SaMD）③ 为基于 AI 的医疗软件提供了风险分类框架；欧盟的 GAIA-X 项目建立了跨境医疗数据共享基础设施，实现了数据主权保护以及 AI 模型的协作

① 参见《2023 年全球人工智能治理指数报告》。

② 参见 Brundage，M.，et al.，*The Malicious Use of Artificial Intelligence：Forecasting，Prevention，and Mitigation*，https://arxiv.org/abs/1802.07228。

③ 参见 https://premierresearch.cn/。

开发，德国的癌症研究中心曾通过这一项目联合法国、意大利训练 AI 肿瘤筛查模型。在自动驾驶方面，联合国欧洲经济委员会通过了全球首个自动驾驶国际法规，涵盖自动车道保持系统、数据记录器等行业标准，欧盟、日本、韩国等 58 国参与其中，中国也于 2021 年加入。中德智能网联汽车合作自 2018—2023 年进行了为期 5 年的试点合作，建立了“上海-慕尼黑测试走廊”，验证跨品牌车辆协同感知技术，构建了数据共享平台，实验覆盖了中国复杂城区与德国高速等场景，华为、奥迪、同济大学等机构也参与了该合作。

（三）全球人工智能治理政策分析专业论坛是有效存在的

联合国系统在人工智能全球治理领域的努力值得被高度关注，特别是其设立的两个重要平台——国际电信联盟（ITU）和联合国区域间犯罪与司法研究所（UNICRI）的人工智能和机器人中心，它们各自在推动人工智能治理方面发挥着不可替代的作用。国际电信联盟（ITU）曾多次举办具有全球影响力的人工智能峰会，为政府、企业和学术界提供了一个中立且重要的交流平台。这些峰会不仅促进了各方对人工智能技术的深入理解，还就标准化、政策制定等关键问题展开了广泛而深入的讨论，为未来的治理实践奠定了坚实的基础。联合国区域间犯罪与司法研究所（UNICRI）的人工智能和机器人中心则专注于在机器人学和人工智能治理的特定领域发挥作用。该中心通过优化协调、知识搜集传播和宣传活动，努力提升公众对机器人学和人工智能技术的认识，促进公众深入理解其优缺点。此外，该中心还提供国际资源，支持各国在机器人学和人工智能的相关事务上开展合作与交流。

非营利组织是全球人工智能治理的重要组成部分。未来生命研究所（FLI）致力于研究如何引导变革性技术造福生命，该研究所通过政策制定与宣传，努力弥合了解变革性技术的专家与具有合法性和治理手段的公共机构之间的差距。未来生命研究所曾发起“暂停 AI 研究”的公开信，呼吁所有 AI 实验室至少在半年内暂停训练比 GPT-4 更强大的 AI 系统，以评估其潜在风险，这一倡议得到了包括马斯克在内的超过 1100 名科技人士的签名支持。电气电子工程师学会（IEEE）作为全球最大的专业技术人员协会，在世界上具有较高的影响力，该机构通过的行业标准是会员必须遵守的规则。IEEE 曾制定多项生物识别、自动驾驶、无人机等领域的行业标准，获得了世界各国的普遍认可。

五、全球人工智能治理的政策学习机制

全球人工智能治理的对话不应沦为少数技术强国和科技巨头的“一言堂”，

而应成为涵盖各国政府、国际组织、科技企业、学术机构、公众等多元主体共同参与的“交响乐”，应该分层次构建合适的学习机制，以应对这一全球性问题。全球人工智能治理跨联盟政策学习网络模型如图 2 所示。

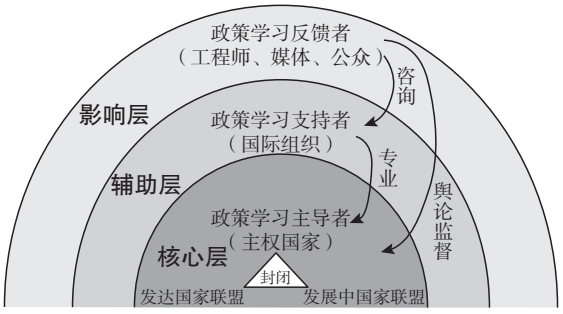

图 2　全球人工智能治理跨联盟政策学习网络模型

资料来源：作者自制。

（一）核心层——构筑稳固的履约合作机制

尽管全球化深刻重塑了全球治理格局，大国间的竞争逻辑依旧稳固不变，国际政治权力结构继续成为塑造全球人工智能治理规范的关键因素。在面对治理挑战时，主权国家作为核心参与者，始终掌握着关键决策权：从制定数字贸易规则到建设风险管控机制，从搭建协同治理框架到国内立法保障，国家层面的主导作用无可替代。同样，在模型训练和技术研发等前沿领域，国家提供的政策环境、资金投入及人才保障是不可或缺的基础。作为全球人工智能治理的核心力量，大国的治理能力直接影响国际规则的执行效果。鉴于中美两国在全球战略地位上的独特性和地域多样性，双方在人工智能治理方面存在广泛的相互学习空间。通过区域治理对比、技术高地对接以及湾区协同创新等方式，双方或许可以创建中美治理对话的实践平台，探索在跨境数据流动、自动驾驶等领域的监管协作，打造具有国际示范效应的治理接口，促进技术领先国家之间在资源整合与人才合作方面的深度联合，为开发符合社会价值导向的 AI 模型奠定基础。这种基于地域特征和技术发展水平的对标研究，既确保了政策创新与本土需求的高度适配，又实现了国际治理经验的协同整合，从而为构建包容、高效的全球人工智能治理体系提供关键的实践支撑（杨超，2025）。

（二）辅助层——构建高效的协调反馈机制

在全球人工智能治理的复杂格局中，亟须构建一个整合跨地域、跨学科、多层次的协同治理机制，以实现全球人工智能治理的敏捷性和多层次性。中国工程院外籍院士张亚勤指出，数据是 AI 深度学习的“石油”，确保数据安全直

接关系到技术的透明性、可靠性和安全性①。首先，构建数据监测与预警系统是基础性模块，该系统的建设包含两个关键环节：其一，实时数据采集与共享，可以通过分布式传感器网络和区块链等技术，实现全球范围内的数据实时流通；其二，智能风险识别，可以运用深度学习、自适应神经网络等机器学习技术，自动检测数据多样性不足、代表性偏差、隐私泄露等风险。其次，构建立体化反馈网络，确保治理的及时性、灵活性与协同性，包括纵向治理通道和横向协同网络两大部分。纵向治理通道可以建立从技术执行层到政策决策层的信息直通车，设立多级联动的 AI 风险管理委员会（地方→国家→国际）；横向协同网络需要构建政府、国际组织、企业、学界和公众之间的信息交互平台，设立科技伦理委员会、公众咨询委员会等机构，整合多元主体的治理智慧，以解决集体行动的成本和效率问题。这种兼具敏捷响应能力和立体治理结构的机制设计，既能够快速应对技术风险，又能确保治理决策的包容性和科学性，为全球人工智能治理提供系统性解决方案（陆卫明，2024）。

（三）影响层——建立交互式传播扩散机制

影响层是政策学习的社会支撑体系。影响层作为政策学习的重要保障机制，由专业共同体、新闻媒体、公众三大主体构成协同作用的支撑网。专业共同体负责知识生产与技术解译，科学家与工程师群体构成核心智库，其作用体现在三大层面，在基础研究层面，通过技术实验揭示大模型运行机理与性能边界；在知识传播层面，借助学术报告、专业著作、国际会议等多渠道输出研究成果；在政策支持层面，为治理谈判提供技术参数与风险评估框架，强化国际合作的科学基础。新闻媒体作为信息中介枢纽，通过议程设置与认知塑造双重机制发挥作用。一方面，新闻媒体进行深度报道和评论解读，培育公众的技术治理素养；另一方面，激发社会讨论，形成政策优化的外部压力。公众参与可以保证治理的合法性，良好的科学素养、多元的参与渠道是公众参与治理取得良好效果的重要保证。这三重主体形成的“知识—传播—参与”闭环，构建了政策学习的动态支撑体系。该体系通过专业知识供给、公共舆论塑造、社会监督反馈的三重机制，既降低了政策制定的信息不对称，又增强了治理方案的社会接受度，最终提升全球人工智能治理的效能与合法性。

六、结论

全球人工智能治理远非简单的技术规制问题。当前国际讨论虽以和平与发

① 参见对张亚勤院士的访谈内容：https://36kr.com/p/2003677951627138。

展为表层叙事，实则关涉两大深层矛盾：一是全球经济利益分配机制的调整，二是人类文明伦理框架的重构。作为非中立性技术的典型代表，人工智能已成为大国间价值理念竞争的新型战略载体。在技术迭代持续加速的背景下，全球人工智能治理主导权的争夺已上升为主要经济体的核心战略议题。各主权国家基于技术主权诉求，相继构建差异化治理体系，但技术风险的不可预测性与地缘政治的博弈产生共振效应，最终导致国际人工智能治理生态呈现“规则碎片化”与“价值阵营化”的双重分裂态势。本研究基于倡议联盟框架（ACF），对全球人工智能治理格局进行了系统性解构。研究发现，当前国际人工智能治理体系主要呈现两种范式分野，“强监管联盟”倾向于构建具有法律约束力的刚性治理架构，主张通过制度性约束机制防范系统性风险；而“弱监管联盟”则倡导非强制性的弹性治理模式，依托市场自律与技术自治，强调通过监管沙盒等创新机制实现安全与发展的动态平衡。进一步分析表明，两大联盟在技术研发、伦理认知、国际规则制定、话语权争夺以及跨境管辖适用性等维度存在根本性的价值分歧，这些结构性矛盾导致全球人工智能治理体系面临持续性的协调困境。研究显示，当前人工智能治理联盟间的互动已满足政策学习的三个基本条件。第一，“智能向善”的伦理共识为技术发展提供了文明基准线；第二，多层治理体系已形成多元化实践成果；第三，联合国等国际组织与非政府机构搭建了专业化的政策对话平台。在机制设计层面，本文建议构建三级协同的政策学习体系：（1）核心层建立制度化的履约合作机制；（2）辅助层完善动态协调与反馈机制；（3）影响层发展参与式传播扩散机制。这三个层级形成闭环系统，通过相互强化实现三重治理效能，确保技术发展的人类利益导向性，提升治理方案的社会适应性，最终实现“科技向善”的价值闭环。

参考文献

阿扎尔（2024）. 指数型技术重塑世界. 郭英、袁冬华，译. 杭州：浙江科学技术出版社.

Hazard.（2024）. *Exponential Technology Reshapes the World.*（Guo Y. & Yuan D. H., Trans.）. Hangzhou: Zhejiang Science and Technology Press.（in Chinese）

保罗·A.萨巴蒂尔、汉克·C.詹金斯（2011）. 政策变迁和学习：一种倡议联盟途径. 邓证，译. 北京：北京大学出版社.

Sabatier, P. A. & Jenkins, H. C.（2011）. *Policy Change and Learning: An Advocacy Alliance Approach.*（Deng Z., Trans.）. Beijing: Peking University Press.（in Chinese）

毕超（2024）. 全球人工智能治理共享共治的理念、体系和机制路径研究. 中国网络空间安全协会人工智能安全治理专委会. 2024 世界智能产业博览会人工智能安全治理主题论坛论文集. 26-29.

Bi, C.（2024）Research on the Concept, System, and Mechanism Path of Shared Governance in Global Artificial Intelligence Governance Chinese Cyberspace Security Association Artificial Intelligence Security Governance Special Committee. Proceedings of the 2024 World Intelligent Industry Expo Artificial Intelligence Security Governance Theme Forum. 26-29.

布莱恩·克里斯汀（2023）. 人机对齐. 唐璐，译. 长沙：湖南科技出版社.

Kristin, B. (2023). *Human Machine Alignment*. (Tang L., Trans.). Changsha: Hunan Science and Technology Press. (in Chinese)

傅莹 (2020). 人工智能的治理和国际机制的关键要素. 人民论坛, (4): 6-8.

Fu, Y. (2020). Key Elements of Governance and International Mechanisms for Artificial Intelligence. *People's Tribune*, (4): 6-8. (in Chinese)

黄锫 (2025). 论我国人工智能领域包容审慎监管的法治维度. 财经法学, (2): 94-109.

Huang, P. (2025). On the Rule of Law Dimension of Inclusive Prudential Regulation in China's AI Field. *Law and Economy*, (2): 94-109. (in Chinese)

何宇、张建华、陈珍珍 (2020). 贸易冲突与合作: 基于全球价值链的解释. 中国工业经济, (3): 24-43.

He, Y., Zhang, J. H. & Chen, Z. Z. (2020). Trade Conflict and Cooperation: An Interpretation Based on Global Value Chains. *China Industrial Economics*, (3): 24-43. (in Chinese)

刘特、郑跃平、曹梦冰 (2024). 算法行政: 文献述评与研究展望. 公共行政评论, 17(1): 84-104+198.

Liu, T., Zheng, Y. P. & Cao, M. B. (2024). Algorithmic Administration: Literature Review and Research Prospects. *Journal of Public Administration*, 17(1): 84-104+198. (in Chinese)

李飞飞 (2024). 我看见的世界: 李飞飞自传. 赵灿, 译. 北京: 中信出版社.

Li, F. F. (2024). *The World I See: Autobiography of Li Feifei*. (Zhao C. Trans.). Beijing: CITIC Press. (in Chinese)

陆卫明、张文、王明 (2025). 全球人工智能治理从"共识"到"共治": 动力、挑战与路径. 西安交通大学学报(社会科学版), 45(2): 1-20.

Lu, W. M., Zhang, W. & Wang, M. (2025). Global Artificial Intelligence Governance from "Consensus" to "Co-governance": Motivation, Challenges, and Paths. *Journal of Xi'an Jiaotong University* (*Social Sciences Edition*). 45(2): 1-20. (in Chinese)

贾开、俞晗之、薛澜 (2024). 人工智能全球治理新阶段的特征、赤字与改革方向. 国际论坛, 26(3): 62-78+157-158.

Jia, K., Yu, H. Z. & Xue, L. (2024). A New Phase in the Global Governance of Artificial Intelligence: Characteristics, Deficits and the Direction of Future Reform. *International Forum*, 26(3): 62-78+157-158. (in Chinese)

沈伟、冯硕 (2022). 全球主义抑或本地主义: 全球数据治理规则的分歧、博弈与协调. 苏州大学学报(法学版), 9(3): 34-47.

Shen, W. & Feng, S. (2022). Globalism or Localism: Divergence, Game, and Coordination of Global Data Governance Rules. *Journal of Soochow University* (*Law Edition*), 9(3): 34-47. (in Chinese)

孙志伟、殷浩铖 (2025). 人工智能时代数字巨头的技术权力及其对"全球南方"的挑战. 国际安全研究, 43(2): 142-164+168.

Sun, Z. W. & Yin, H. C. (2025). Technological Power of Digital Giants in the Age of Artificial Intelligence and Its Challenge to the Global South. *Journal of International Security Studies*, 43 (2): 142-164+168. (in Chinese)

王海滨 (2025). 美国涉华数据出境限制政策评析. 现代国际关系, (1): 105-124+149.

Wang, H. B. (2025). Review of the US' Restriction Policy of Data Outbound Transfer Towards China. *Contemporary International Relations*, (1): 105-124+149. (in Chinese)

王威力、王玥珺 (2025). 全球 AI 治理中的社会技术想象——人工智能发展规划的国际比较分析. 情报资料工作, 46(3): 1-32..

Wang, W. L. & Wang, Y. J. (2025). Socialtechnical Imaginaries in Global AI Governance: International Comparative Analysis of Artifical Intelligence Development Plans. *Information and Documentation Services*, 46(3): 1-32. (in Chinese)

王彦雨、李正风、高芳 (2024). 欧美人工智能治理模式比较研究. 科学学研究, 42(3): 460-468+491.

Wang, Y. Y., Li, Z. F. & Gao, F. (2024). A Comparative Study of Europran and American AI Governance Models.

Studies in Science of Science, 42(3): 460-468+491. (in Chinese)

徐菲、石斌(2024). 技术民族主义及其对美国对华科技政策的塑造. 当代美国评论, 8(2): 60-81+127-128.
Xu, F. & Shi, B. (2024). Techno-nationalism and Its Influence on U. S. Technology Policy Towards China. *Contemporary American Review*, 8(2): 60-81+127-128. (in Chinese)

杨超、乔瀚、李如烟(2025). 美国人工智能监管政策的复杂性及临近性研究——基于州域层面政策文本的主题模型. 科学学研究, 2025. 3. 10, 1-18. https://doi. org/10. 16192/j. cnki. 1003-2053. 20250307. 005 .
Yang, C., Qiao, H. & Li, R. Y. (2025). The Complexity of U. S. Artificial Intelligence Regulatory Policy and Sino-U. S. Regulation Policy Similarity: Based on Topic Model Analysis of State-Level Policy Texts of the U. S. . *Studies in Science of Science*. 2025. 3. 10, 1-18. https://doi. org/10. 16192/j. cnki. 1003-2053. 20250307. 005. (in Chinese)

杨胜刚、李川(2025). 大模型时代数据跨境隐性国家安全风险的敏捷治理模式. 情报杂志, 2025. 5. 15, 1-8. http://kns. cnki. net/kcms/detail/61. 1167. G3. 20250514. 1618. 006. html.
Yang, S. G. & Li, C. (2025) An Agile Governnance Framework for Hidden National Security Risks of Cross-Border Data in the Age of Big Models. *Intelligence Magazine*, 2025. 5. 15, 1-8. http://kns. cnki. net/kcms/detail/61. 1167. G3. 20250514. 1618. 006. html. (in Chinese)

张东冬(2025). 数字外交强化与美国全球数字竞争的全新图景. 国际关系研究, (1): 35-56+156.
Zhang, D. D. (2025). The Strengthening of Digital Diplomacy and U. S. Global Digital Competition New Picture. *Journal of International Relations*, (1): 35-56+156. (in Chinese)

张亮、陈希聪(2025). 生成式人工智能背景下的跨境数据安全规制——基于 DeepSeek、ChatGPT 等主流 AI 的思考. 湖北大学学报(哲学社会科学版), 52(2): 120-128+199.
Zhang, L. & Chen, X. C. (2025). Analysis of Cross-border Data Security Regulation in the Generative AI Context: Based on the Application of Mainstream AI Models such as DeepSeek and ChatGPT. *Journal of Hubei University (Philosophy and Social Sciences)*, 52(2): 120-128+199. (in Chinese)

朱明婷、徐崇利(2023). 人工智能伦理的国际软法之治: 现状、挑战与对策. 中国科学院院刊, 38(7): 1037-1049.
Zhu, M. T. & Xu, C. L. (2023). International Soft Law Governance of Artificial Intelligence Ethics : Current Situation, Challenges and Countermeasures. *Bulletin of Chinese Academy of Sciences*, 38(7): 1037-1049. (in Chinese)

Bradford, A. (2020). *Brussels Effect*: *How the European Union Rules the World*? New York : OxfordUniversity Press.

Dafoe, A. (2018). *AI Governance*: *A Research Agenda*. Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK.

Yudkowsky. & Eliezer. (2008). Artificial Intelligence as a Positive and Negative Factor in Global Risk. *In Global catastrophic risks*. New York: Oxford University Press, 308-345.

责任编辑: 郑跃平

英文目录与摘要

JPA Journal of Public Administration, Vol. 18 No. 3, 2025

●SYSPOSIUM: Policy Implementation Study

Policy Implementation Strategies under Certainty and Uncertainty: A Study on Discretionary Decision-Making in Public Policy ····· Shenghui Li , Dan Li & Shaotang Zhu

Abstract Examining the discretionary strategy selection of implementers—who serve dual roles as both loyal agents and local interest representatives under conditions of certainty and uncertainty helps unpack the "black box" of policy implementation and analyze their behavioral logic. By leveraging the discretion shaped by certainty and uncertainty, along with considering the pressure of policy performance evaluations, this study constructs a "power-pressure" model. Within this model, implementers adopt four strategies: fluctuating implementation, symbolic implementation, political implementation, and administrative implementation, leading to four corresponding outcomes: short-term effectiveness, poor effectiveness, significant effectiveness, and mixed effectiveness. Through a comparative multi-case analysis of four major ecological conservation policies in China, the study reveals the mechanism behind implementers' discretionary strategy selection in contexts of certainty and uncertainty.

Key Words Certainty; Uncertainty; Differential Coping; Implementation Strategy

The Logic of Local Social Policy Implementation: The Case of "Multi-School Zoning" Reform for Compulsory Education ····· Yu Zeng & Guangming Jiang

Abstract In recent years, China has formulated and implemented reforms such as the "Multi-School Zoning" policy to promote educational equity. This study examines local variations in policy implementation mainly from four theoretical perspectives, including administrative pressure, public demand, career incentives for officials, and developmental social policies. It reveals that local policy implementation simultaneously pursues two primary objectives: equity and development. By coding the compulsory education policies of 283 prefecture-level administrative regions and conducting event history analyses, this study finds that when the growth of land revenue slows down, local governments tend to promote equalization of compulsory education resources. This conclusion reflects a close relationship between the implementation of educational equity policies and local economic development. The findings provide valuable theoretical insights and empirical references for improving equal access to basic public services and advancing high-quality economic and social development.

Key Words Education Equity; Developmental Social Policy; Multi-School Zoning; Compulsory Education; Equalization of Basic Public Services

●ARTICLES

How Can the Special Strike Campaign of Grassroots Digital Formalism Be Effective? An Analysis Based on the Campaign Initiators ····· Jingyu Zhang, Tian Gan & Yong Guo

Abstract Formalism is a persistent challenge in grassroots governance, and the issue of digital

formalism brought about by digital transformation has become increasingly prominent. Local governments have attempted to address this problem through special strike campaigns. However, under the current administrative system, can these campaigns achieve long-term effectiveness? Existing studies generally view special strike campaigns as a specific manifestation of campaign-style governance. Due to issues such as collusion, high mobilization costs, and disruption of the normal functioning of bureaucratic organizations, the effectiveness of such measures remains questionable. Nevertheless, City A in Zhejiang Province established a dedicated task force led by the Party Office, the Commission for Discipline Inspection, and the Big Data Bureau to tackle "formalism at the fingertips," achieving notable results and maintaining them over an extended period. This provides us with an insightful case study. Unlike previous research, this study focuses on the initiators of special strike campaigns and finds that having the initiating departments and the rectified departments belong to different systems is key to solving the issue of collusion. Furthermore, if the initiating departments possess a rational power structure, this kind of special strike campaigns can effectively carry out thorough investigations to obtain genuine information, engage in flexible consultations to formulate scientific plans, employ reverse mobilization to overcome implementation difficulties, reduce collusion and enhance supervision quality, and ultimately establish long-term mechanisms. This approach performs well across five procedural elements—information, planning, execution, supervision, and conclusion—achieving desirable rectification outcomes. This finding not only expands theoretical understanding of campaign-style governance, but also provides practical insights for effectively addressing grassroots formalism and improving grassroots governance performance.
Key Words Grassroots Formalism; Digital Transformation; Special Strike Campaingn; Campaign-style Governance

Understanding the Complexity and Rationale of Agile Transformation in Urban Governance: A Research Perspective Based on Deconstructing Institutional Action Situations

..... Tingjia Chen, Ziyi Zhao & Tao Sun

Abstract Agile governance provides new mindsets and alternatives for urban governance in the digital age. Aiming at conceptual development, existing research focuses on normative discussions of agile governance with illustrative cases, and develops a set of agile principles. However, urban governance system is complex and multifaceted. Current research struggled to provide generalizable theoretical explanations and rigorous empirical analyses for the complexity. This study brings local public economy theory to agile governance research, and examines the rationale of urban agile transformation by employing action situation as the unit of analysis. The authors conducted a three-year continuous fieldwork on the "Holistic Smart Urban Governance" practice of J District in W municipality. The study finds that, firstly, agility varies in both quality and frequency across different levels of action situations. The contents of agile transformation at various action situations are different. As the level of action situation deepens, agility diminishes, and stability becomes prominent. Secondly, the interaction among action situations determines the overall effectiveness of the agile practice. The incompatible local institutions constrain the agile transformation of local public service production and provision. Thirdly, non-governmental entities and Information and Communication Technology tools play a significant role in promoting agile transformation at the lower-level action situations, even in the absence of intentional comprehensive design and agile mindsets. However, as the level of action situation deepens, rational

choices and proactive actions by the government become crucial. This study advances the theoretical exploration and empirical examination of agile governance in academia, and provides policy implications for spurring agile transformation in practice.
Key Words Agile Governance; Urban Governance; Action Situation; Public Service

How Do Historical Legacy Issues Arise in Grassroots Governance? A Case Analysis of the "Presentism" Phenomenon in Government Decision-making ····· Chenglong Yu & Xiaohua Zheng
Abstract Historical legacy issues in grassroots governance are not unresolved due to prolonged waiting, but rather because the "policy window" has not been further opened to the "decision-making window." The "presentism" strategy in government decision-making is the deep-seated cause of such problems. The case of Coastal Village Q demonstrates that early officials achieved accelerated development through "active presentism," stacking current resources and compressing current costs, yet left problems for future generations. Successive officials regarded these legacy issues as social risks and addressed them promptly through "passive presentism," such as delaying, escalating, transferring, or patching up the problems, which also led to their continuous postponement. Consequently, the continuous decision-making continuum of "presentism" by local governments at different times has made problems that should have been resolved initially increasingly difficult to solve as institutional and social environments evolve. In fact, under the time pressure brought by development opportunities and social risks, local governments' choice of "presentism" as a coping strategy is largely situational rationality. The paradox lies in that the external conditions granting it rationality will inversely weaken its rationality in the future, thereby leading to the continuous reproduction of historical legacy issues in grassroots governance.
Key Words Historical Legacy Issues; Grassroots Government; Decision-making; "Presentism"; Time

Activating "Public" and "Common": County-Level Social Organizations Bridging Structural Holes ····· Xuefei Lin & Yajie Shan
Abstract The rural livelihood supply network typically involves both the government-led public sector and the locally rooted common sector. Both sectors operate as closed systems, independently and in parallel, relying on differentiated organizations and resources. The Y Foundation in Xiong Town occupies the structural hole between the two sectors, serving as a bridging mechanism. It integrates existing resources from both public and common sectors through strategic alliances, joint decision-making processes, and other collaborative methodologies. The Y Foundation in Xiong Town leverages the charity organization model to conduct volunteer activities and deliver specialized services, thereby driving innovative production in livelihood supply. Through government-society cooperation, it effectively enhances the quality and expands the coverage of local livelihood provision systems. The bridging role of social organizations in rural areas is driven not only by the rational considerations of local governments to enhance their foundational power but also by the societal demand for the modernization of rural philanthropy, rather than being a result of a strong side taking over the other side. This relationship between government and society stems from the clan social network, the philanthropic culture, and stable government-business connections in rural areas. Together, these three elements form

a balanced distribution of political and social power and facilitate government-society elites to engage in collective action based on local moral values and norms. Therefore, research on social organizations anchored in county-level societies can uncover indigenous characteristics of government-society interactions and innovative institutional pathways therein.

Key Words Livelihood Supply Network; Bridging Intermediary; Government-society Interaction; County-level Space; Rural Society

Information Intervention and the Willingness of Flexible Workers' to Participate in Social Pension: A Randomized Controlled Experiment ..... Menghan Shen, Chuping Su & Guangjun Shen

Abstract Flexible workers constitute a significant part of the labor force. Current policies related to the participation of these workers in social pension primarily encourage voluntary participation. However, in practice, a large portion of flexible workers are not enrolled in social pension. This paper employs randomized experimental methods to study the impact of information intervention on the willingness of workers in food delivery couriers to take up social pension. The study finds that providing information about the benefits of pension insurance significantly increases couriers' willingness to participate, as well as their willingness to extend the duration of their participation and increase their contributions. The increase in willingness is more pronounced among groups with higher education levels, higher incomes, older psychological ages, those planning to work as couriers long-term, and those with higher financial literacy. Additionally, information intervention reduces the preference for having male children among these workers, potentially helping to address the imbalance in the gender ratio in China.

Key Words Information Intervention; Social Pension; Take-up; Flexible Workers; Randomized Control Experiment

The Inequality in Municipal Governments' Cooperation: A Study Based on the Network of Offsite Visits by Municipal Officials in the Yangtze River Delta

..... Gangjian Lin, Yuanshuo Xu & Weiwen Zhang

Abstract The high-quality development of urban agglomerations relies on a balanced and reciprocal network of municipal cooperation. However, existing studies mostly simplify municipal cooperation to undirected and equal connections, overlooking the differences in cooperation willingness and resource investment, namely the cooperation inequality. This study uses large language models to identify 16,039 cooperative offsite visit events from over 720,000 government website news data, constructing a panoramic network of official offsite visits among 41 cities in the Yangtze River Delta from 2010–2023. It analyzes the characteristics and influencing factors of inequality in informal cooperation among municipal governments. The study finds that the overall cooperation inequality in the Yangtze River Delta shows a declining trend, with heterogeneity across different cooperation fields and levels. At the city level, both efficiency logic and administrative logic jointly determine the cooperation strategies between municipal governments, affecting cooperation inequality between different cities and forming complex cooperation network characteristics. The study innovatively uses large language model technology for empirical research on local government networks and conducts an empirical analysis of bilateral cooperation inequality for the first time, providing a basis for understanding the current state

and future path of regional cooperation development in urban agglomerations.
Key Words Offsite Visits; Large Language Model; Local Government Network; Yangtze River Delta; Cooperation Inequality

Alliance Conflict and Policy Learning in Global Artificial Intelligence Governance: Based on the Advocacy Coalition Framework ····· Xiaokang Lei & Linxia Yu
Abstract Global artificial intelligence governance has gone beyond the scope of simple technical regulation and evolved into a deep game involving the ethical foundation of human civilization and the reconstruction of international political and economic order. As a strategic technology with significant non neutrality characteristics, AI is becoming a new carrier of value system competition among major powers, and the struggle for governance dominance has become a strategic priority for major economies. However, the uncertainty of technological iteration combined with geopolitical competition has led to a dual alienation of the international governance ecosystem, characterized by fragmented rules and value camps. There are fundamental differences between the "strong regulatory alliance" and the "weak regulatory alliance" in terms of technical ethics, rule making power, and cross-border jurisdiction. The existing governance mechanism is difficult to resolve the coordination dilemma caused by these structural contradictions. However, the ethical paradigm of "intelligence for good", multi-level governance practices, and professional dialogue platforms constructed by international organizations provide the possibility to break through the governance deadlock. Based on the initiative alliance framework, it is suggested to establish a three-level collaborative mechanism: the core layer strengthens institutional compliance, the auxiliary layer optimizes dynamic coordination, and the influence layer improves participatory communication. This three-dimensional design can ensure the human centered orientation of technological development, enhance the regional adaptability of governance plans, and ultimately achieve a positive cycle of technological ethics and human values.
Key Words Artificial Intelligence Governance; Advocacy Coalition Framework; Strong Regulatory Alliance; Weak Regulatory Alliance; Policy Learning

●THEORETICAL REVIEWS

Meta-Analysis in Public Administration: Progress, Paradigms, and Directions
····· Qing Miao & Xingkui Zhu
Abstract As a powerful methodological tool for reconciling inconsistent research findings, meta-analysis is exerting a growing influence on the paradigmatic development of public administration and the broader social sciences. First, this study provides a concise overview of the historical evolution and distinct advantages of meta-analytic methods. Second, it systematically reviews 54 meta-analyses published in leading domestic and international public administration journals over the past two decades, mapping their publication trends and thematic focuses. Third, it summarizes ten essential operational steps for conducting rigorous meta-analyses. Finally, it introduces three advanced meta-analytic approaches, including three-level meta-analysis, and outlines their potential to foster theoretical and methodological innovation in public administration research.
Key Words Meta-Analysis; Research Paradigm; Public Administration; Methodological Innovation

双月刊，2008年创刊
第18卷，第3期（总第105期）
2025年6月15日出版

Journal of Public Administration
Bimonthly, Since 2008
Vol.18 No.3
Published in June 2025

公共行政评论

主管单位 中华人民共和国教育部
主办单位 中山大学
协办单位 教育部人文社会科学重点研究基地-
中山大学中国公共管理研究中心
广东省行政管理学会
社　　长 肖　滨
主　　编 朱亚鹏
联系电话 020-84113029　020-84038746　　传　　真 020-84111478
电子邮件 jpachina@163.com　　网　　址 http://jpa.sysu.edu.cn
编辑出版 《公共行政评论》编辑部
（广州新港西路135号；邮编510275）
印　　刷 广州一龙印刷有限公司
国内发行 广东省报刊发行局
国外发行 中国国际图书贸易总公司
（北京399信箱）

Administrator Ministry of Education of the People's Republic of China
Sponsors Sun Yat-sen University
Supporters Centre for Chinese Public Administration Research，Sun Yat-sen University
Guangdong Public Administration Society
President： Bin Xiao
Editor in Chief： Yapeng Zhu
Tel：86 20 84113029　86 20 84038746　　Fax：86 20 84111478
Email：jpachina@163.com　　Website：http://jpa.sysu.edu.cn
Edited by Editorial Office of *Journal of Public Administration*
(NO.135 Xin Gang Xi Road,Guangzhou,China.510275)
Distributed by China International Book Trading Corporation
(P.O.Box 399,Beijing,China)

刊号 ISSN1674-2486 CN44-1648/D
邮发代号 国内46-364 国外BM8839
国内外公开发行
国内定价：20.00元

ISSN 1674-2486