战略决策研究 （双月刊）

JOURNAL OF STRATEGY AND DECISION-MAKING 2024年 第3期（总第86期） 第15卷

目 录

期刊基本参数：CN44-1673/C*2010*B*A4*102*ZH*P*￥15.00*1000*05*2024-05

Journal of Strategy and Decision-Making (Bimonthly)

Table of Contents

发展新质生产力与共建“一带一路”双向赋能：现实逻辑、理论框架与实践进路

韩永辉　李思怡　成　昊

摘　要：当前，世界进入新的动荡变革期，单边主义思潮再度涌现，发展新质生产力和共建“一带一路”高质量发展面临的外部风险更为突出。为发挥多边力量、应对外部风险，本文首先厘清发展新质生产力与共建“一带一路”间的内涵联系，进而分析二者间存在的互促发展机制、构筑二者间双向赋能的理论框架，并基于此探寻发展新质生产力与共建“一带一路”双向赋能的实践路径。

关键词：新质生产力战略；“一带一路”倡议；高质量发展

作者简介：韩永辉，广东外语外贸大学广东国际战略研究院、中共广东省委党校2024春季中青年干部培训三班、广东省习近平新时代中国特色社会主义思想研究中心广东外语外贸大学基地，教授；
李思怡，广东外语外贸大学高级翻译学院科研助理；
成昊，广东外语外贸大学广东国际战略研究院硕士研究生，本文通讯作者，E-mail：anroycheng@163.com。

一、引言

目前，全球格局正处于复杂动荡的变革期，世界经济复苏未达预期；国际力量对比的深刻调整使得单边主义、保护主义、霸权主义思潮再度涌

本文系国家社会科学基金重大项目（项目编号：21&ZD074）、国家自然科学基金资助项目（项目编号：71873041；72073037）、广东省自然科学基金项目（项目编号：2022B1515020008）、广东外语外贸大学全球治理与人类命运共同体重点实验室和广东省哲学社会科学创新工程2022年度特别委托项目（项目批准号GD22TWCXGC12）的阶段成果。感谢匿名评审专家和编辑部对本文提出的意见和建议，文责自负。

现，全球各国外源风险水平不断提高，显著影响多边合作机制的深化发展。基于中国视角，外部环境动荡已成为中国经济向好发展的主要挑战之一，产业新旧动能转换、市场资源错配、劳动力供需结构失衡等问题亦一定程度上影响高质量发展进程。[①] 基于“一带一路”视角，共建“一带一路”历经十年发展，已取得显著成效，但共建国家当中的发展中国家占比较高，整体经济、技术和基建水平相对落后，[②] 产业数字化发展基础亦较为薄弱，[③] 仍面临较为严重的全球价值链“低端锁定”问题。

面对全球动荡变革时局和新一轮技术革命、国内经济转型发展等关键时点的历史性交汇，习近平总书记于2023年9月在黑龙江考察期间指出要“整合科技创新资源，引领发展战略性新兴产业和未来产业，加快形成新质生产力”，首次提出了“新质生产力”这一概念。新质生产力概念既是中国对马克思主义生产力理论的又一重大创新，也是中国妥善应对百年未有之大变局下的外部风险、紧抓国际格局演变重构所携机遇的关键思想引领。从新质生产力的内涵来看，其“新”所代表的关键性颠覆性技术突破和“质”所代表的创新驱动本质，[④] 不仅将更好推动中国高质量发展进程，还可依托“一带一路”框架向外产生更大正向外溢。为此，本文在准确把握新质生产力内涵及特征的理论基础上，厘清发展新质生产力与推动共建“一带一路”高质量发展间的互促机理，探寻发展新质生产力与共建“一带一路”协同并进的现实路径，为中国统筹国内国际两个大局、持续推动高水平对外开放、加快社会主义现代化强国建设步伐提供研究助力。

二、新质生产力赋能共建“一带一路”高质量发展的作用机制

新质生产力作为一种新型生产力形态，其内在属性呈现出鲜明的人民

① 贺颖、倪红福等：《高质量发展背景下中国经济面临的重大问题及对策建议》，载《财经智库》2023年第8期，第47-70、145-146页。

② 屠年松、郑雅哲等：《“一带一路”倡议与全球价值链升级——基于沿线国家的实证数据》，载《经济问题》2024年第5期，第52-60页。

③ 李青、易爱娜等：《“一带一路”数字合作：成就、挑战与展望》，载《战略决策研究》2023年第6期，第51-67页。

④ 周文、许凌云：《论新质生产力：内涵特征与重要着力点》，载《改革》2023年第10期，第1-13页。

性、协调性与开放性特征。[①] 这与马克思主义理论体系中关于社会发展根本目标的论述高度契合，即新质生产力的培育和发展最终将服务于广大人民群众福祉的显著提升。而“一带一路”倡议同样将推动各国社会全面进步、助力沿线各国人民生活水平普遍改善作为重要目标。可见，发展新质生产力与共建“一带一路”在发展目标上具有内在的高度一致性。随着共建“一带一路”的制度型发展日益完善和多边合作的持续深化，中国在新质生产力培育过程中取得的创新性成果与变革性突破，必将通过“一带一路”合作机制向共建国家产生积极的溢出效应。

（一）以前沿技术突破赋能“一带一路”科技化发展

新质生产力源于颠覆性、根本性技术突破，相较于传统生产力，其对既有技术路径和范式的创新性重塑更为显著。在新质生产力引领下，一系列新产品、新服务的涌现将推动产品架构、商业模式乃至应用场景的加速变革，[②] 由此形成强大内生动力将带动中国技术创新能力的整体跃升，进而突破全球价值链“低端锁定”的发展困境。目前，中国在部分关键技术领域面临的“卡脖子”难题，以及西方发达国家在单边主义思潮影响下对华实施的技术封锁与科技制裁，已成为阻碍中国实现核心技术突破、提升全球价值链地位的重要制约因素。作为技术变革催生的先进生产力形态，新质生产力的发展壮大将伴随一系列战略性新兴产业的迅猛发展、诸多关键核心技术的接续突破以及新旧动能加速转换的宏观格局，最终推动中国全球价值链地位持续攀升。

基于“一带一路”视角，中国在新质生产力培育过程中所取得的技术突破和全球价值链位势攀升成效，可通过“一带一路”合作框架在共建国家中产生大规模的正向外溢效应。一方面，共建“一带一路”为中外技术合作奠定了良好的制度基础。2023年11月召开的首届“一带一路”科技交流大会上，中国携手各共建国家一同提出了《国际科技合作倡议》，该倡议旨在倡导践行开放、公平、公正、非歧视的国际科技合作理念，坚持“科学无国界、惠及全人类”，明确提出“一带一路”共建国家要“携手

① 张森、温军：《数字经济赋能新质生产力：一个分析框架》，载《当代经济管理》2024年第5期，第1–12页。

② 李晓华：《新质生产力的主要特征与形成机制》，载《人民论坛》2023年第21期，第15–17页。

构建全球科技共同体”。[①] 同时，“一带一路”框架下的双边科技合作协定不断增加。至2023年底，中国已和超过80个“一带一路”沿线国家签署政府间科技合作协定。结合中国与各“一带一路”共建国家签署的合作协定及发展规划、谅解备忘录来看，深化技术合作、搭建中外双边技术合作框架和机制，已成为多数“一带一路”共建国家的共识。“一带一路”框架下的中外双边技术合作共研渠道已较为畅通，新质生产力发展所带来的突破性、颠覆性技术成果可通过这一渠道实现良好正向外溢。另一方面，共建“一带一路”合作项目的技术密集度逐步提高。相较于“一带一路”倡议提出初期，近几年来中国对“一带一路”共建国家合作中的技术要素引领作用更为突出，部分基于中国具备领先优势技术所开展的“一带一路”合作项目已取得良好成效。截至2023年末，中国已与各共建国家以合作形式建设了“一带一路”联合实验室50多家、跨国技术转移中心9个，累计举办技术交流对接活动300余场，促进千余项合作项目落地。

目前，中国已依托“一带一路”框架，在东南亚、中东和非洲等地的多个沿线国家开展了众多技术共享和投资合作项目，合作成效已得到全球认可。如中国与印度尼西亚在采矿业及新能源技术领域的共享合作，已带动印尼由粗镍矿出口大国转变为全球最大精炼镍生产国；中国与阿拉伯国家的光伏、风电技术合作亦得到阿拉伯国家广泛认同，合作范围由能源领域不断向5G通信、人工智能领域拓展，并带动了阿拉伯国家对华投资规模的迅速上升。可以预期，中国在加快培育新质生产力过程中所取得的核心关键技术突破将发挥出更大的对外溢出效应，并通过技术共享、项目共建等形式提升“一带一路”沿线国家的技术发展水平，最终推动“一带一路”共建国家逐步脱离全球价值链“低端锁定”困境，实现科技化发展。

（二）以数字经济合作赋能“一带一路”数字化发展

作为先进生产力的具体表现形式，新质生产力具有高科技、高效能、高质量等鲜明特征。一方面，新质生产力与数字技术、数据要素、数字经济的深度融合是其区别于传统生产力的突出表征；另一方面，新质生产力的发展壮大也离不开以数字化转型为重要内容的现代化产业体系建设。数据要素作为重要的创新要素之一，其对市场主体生产效率的提升作用已得

① https://www.gov.cn/yaowen/liebiao/202311/content_6914071.htm.

到验证。[①] 中国政府亦高度强调数字技术、数据要素在发展中的作用，相继出台了《“十四五”数字经济发展规划》《关于构建数据基础制度更好发挥数据要素作用的意见》《“数据要素×”三年行动计划（2024-2026年）》等政策措施。可见，对新质生产力的大力培育，实则是推动产业数字化转型、建设现代化产业体系的深化与延申，在新质生产力培育过程中，数字技术革新、数据要素整合、数字经济培育等数字化举措都将成为关键抓手。

与此同时，“一带一路”沿线国家对数字经济发展的重视程度也在迅速提高，对数字技术、数据要素的需求不断扩大。在2023年10月18日召开的第三届“一带一路”国际合作高峰论坛数字经济高级别论坛上，智利、阿根廷、肯尼亚、阿联酋等多个国家的高级官员出席并致辞，代表各共建国家表达了深化数字领域合作的强烈意愿，并一致推动了《“一带一路”数字经济国际合作北京倡议》《航运贸易数字化与“一带一路”合作创新白皮书》等协议达成。可见，中国加快培育新质生产力，在带动本国数字经济高质量发展、数字技术变革性突破的同时，也将通过“一带一路”合作机制赋能共建国家数字化发展进程。具体而言，一方面，中国数字技术的发展将推动“一带一路”区域经济一体化水平提升。数字经济已成为推动区域经济一体化的关键力量，云计算、大数据和人工智能等数字技术通过优化资源配置和提高生产效率的方式，不仅可促进地区间的经济协同增长，还将增强区域内商品与服务的流通性。[②] 同时，由数字技术突破带来的实时数据分析和信息交换系统的建立，亦有助于共建国家更好对接彼此市场、应对国际经济变化，进一步加强区域互联互通水平。这种数字化发展不仅可加速各成员国的经济现代化进程，也将为“一带一路”合作框架注入新的活力，推动沿线国家数字化转型和经济现代化发展迈上新台阶。另一方面，中国与共建国家开展的数字基建设施共建将直接带动“一带一路”共建国家数字化水平提高。推动高质量数字基础设施的共建是“一带一路”倡议中实现数字化转型的关键策略。中国新质生产力促进的数字技术进步，将提升“一带一路”数字基建合作项目的技术密集度，

① 赵宸宇、王文春等：《数字化转型如何影响企业全要素生产率》，载《财贸经济》2021年第7期，第114-129页。

② 包振山、韩剑等：《数字经济如何促进对外贸易高质量发展》，载《国际经贸探索》2023年第2期，第4-20页。

进而增强共建国家网络连接稳定性和数据处理效率，更有力地支撑区域内的数字经济活动。此外，数字基础设施的现代化改造和互联互通，也将有效助推共建国家数字化转型进程，为“一带一路”数字化、现代化发展注入新动能。

（三）以国际分工优化赋能“一带一路”全球化发展

基于产业视角，新质生产力是企业采用新型要素组合方式、提供新产品以更好满足社会需求，从而在“企业集合”意义上形成新业态和新结构，[①] 具体表现为传统产业向新兴产业，尤其是战略性新兴产业的转变。新质生产力的加快培育将带动中国现代化产业体系构建和产业转型升级发展，助力中国抢占战略性新兴产业及未来产业的全球先机，同时也将倒逼相对落后产业的对外转移。

当前，部分“一带一路”共建国家仍处于经济发展的初级阶段，国内产业水平较低、国际分工参与不足，难以充分发挥自身的资源禀赋和比较优势以在全球市场竞争中取得更大收益。中国新质生产力培育所带来的生产效率提升和产能转移加速，将通过以下三个渠道助力处于相对落后状况的“一带一路”共建国家进一步融入国际分工体系、实现全球化发展。一是，新质生产力可促进中国生产智造水平提升，为“一带一路”共建国家提供更多优质生产工具。中国在采矿及冶炼、农机等领域内的技术水平提升将为共建国家提供更为高效可靠的生产工具，提升其生产制造效率、缩减产品社会必要劳动时长，进而推动其依托自身资源禀赋优势，扩大具有比较优势产品的生产和出口，最终带动其在国际分工中的地位和话语权不断提升。二是，新质生产力发展倒逼中国部分产业产能向外转移，提升相对落后的“一带一路”共建国家的产业体系现代化水平。为更好拓展高技术制造业发展空间、引导产业有序转移，中国政府已于2022年出台《关于促进制造业有序转移的指导意见》。在加快培育新质生产力背景下，部分偏向低端的制造业产能亦因成本上涨等因素存在向外转移意愿。而当前部分“一带一路”共建国家仍未形成成熟的工业体系和产业链条，甚至有部分国家因战乱导致产业体系严重受损，中国向外转移的部分制造业产能恰好可丰富这些国家的产业链环节，并带动当地就业和经济发展。三是，

① 高帆：《“新质生产力”的提出逻辑、多维内涵及时代意义》，载《政治经济学评论》2023年第6期，第127–145页。

新质生产力提升中国终端消费品生产效率，带动“一带一路”共建国家生产专业化水平提高。目前，部分“一带一路”共建国家生产能力相对薄弱，在粮食、医疗器械、电子产品等需求刚性高、规模大的产品生产中效率较低，且由于全球产业链供应链波动加剧，部分国家甚至因物资短缺出现大规模社会问题，政府由此受迫向相关供应链条投入大量资金以提升关键物品供给水平。新质生产力所带来的产业转型升级和生产效率提升恰可以协助“一带一路”共建国家应对此类问题。中国可依托生产效率提升、产业门类齐全的良好基础条件，针对性向“一带一路”共建国家出口其急需物品，以降低部分“一带一路”共建国家产业链供应链压力；同时，亦可缓解共建国家政府财政压力，避免政府因物资极端紧缺而采取短视投资、大量进口等行为，并带动共建国家资本合理配置至本国具有比较优势的产业当中，推动优势产业加快发展、提质增效，在全球市场和国际分工体系中取得更大优势，反哺国内经济发展。

三、“一带一路”高质量发展对新质生产力培育的反馈机制

新质生产力的培育是一个复杂而系统的过程，其发展受到科技创新能力、人才储备质量、产业转型升级进程等诸多内在因素的影响。新质生产力的形成，是在传统生产力发展的基础上引入新科技、新业态、新模式等，进而促进生产力发展范式的转变，[①] 由此，新质生产力的培育将对核心技术、高端人才、关键原材料等战略性新兴产业发展所必须的要素提出更大的需求。然而，当前中国经济仍然面临国外复杂严峻的风险挑战和国内多重因素交织叠加带来的下行压力，有效需求不足、不可再生资源受限、核心城市空间不足等问题亦对新质生产力的培育带来一定阻碍。作为共建“一带一路”的倡议国和主要推动力量，中国应对新质生产力培育所面临的挑战，需充分用好共建“一带一路”所带来的时代机遇，加强与“一带一路”共建国家在更大范围、更广领域、更高水平上的交流合作，依托多边合作带来的外源助力，推动新质生产力加速生成。

① 柳学信、曹成梓等：《大国竞争背景下新质生产力形成的理论逻辑与实现路径》，载《重庆大学学报（社会科学版）》2024年第1期，第145–155页。

（一）以资源供给稳定助力产业链条韧性提升

新质生产力发展与战略性新兴产业和未来产业紧密相关，[①] 前沿产业和技术的发展将在较大程度上影响一国在新质生产力领域内的全球竞争优势。目前，全球各国在前沿产业、关键技术上的竞争日益加剧，新质生产力逐渐成为大国博弈的重要阵地，中国在新质生产力的培育过程中将面临更多制裁和打压。

当前中国部分战略性新兴产业和未来产业链条中，仍有部分关键原料的对外依存度相对较高，产业链供应链所面临的“卡脖子”问题仍然存在，“一带一路”倡议恰可通过其稳定的政策沟通渠道及共建国家种类丰富、总量巨大的资源储备应对这一问题。以中国具备全球领先地位的新能源汽车产业为例，其制作电池正负极所需的关键材料镍和铜的进口依存度均偏高且暂无可行替代原料，一旦出现镍、铜矿进口受阻情况，新能源汽车产业链条将遭受严重冲击甚至生产停滞。同时，海外市场镍矿供应量及价格的波动风险较为显著。2020年，全球镍矿出口第一大国印度尼西亚曾以降低经济对镍矿出口依赖度为由，出台镍矿出口限制措施，导致全球镍价大幅上涨，伦敦有色金属交易所（LME）镍价一度10万美元/吨以上；2024年3月，因印度尼西亚矿产和煤炭开采业务活动的工作计划和预算报告（RKAB）整体发放进度缓慢，市场产生镍矿供应不足的担忧情绪，再度引发镍价在单月内上涨逾12%；加之2024年4月13日英美两国再度对俄施加制裁，迫使芝加哥商品交易所（CME）和伦敦有色金属交易所（LME）禁止接受俄罗斯镍矿，进一步加剧了全球镍矿供应量及价格波动风险。

随着共建“一带一路”的纵深推进，中国与共建国家政策沟通的广度和深度不断拓展，双方通过高层互访、发表联合声明、签署合作协议等方式达成了诸多重要共识。这种密切的政策对话不仅大大增进了彼此之间的政治互信，也进一步夯实了共同发展的行动基础。更为重要的是，以战略对接、优势互补为基础，中国与共建国家携手推动构建起多层次、宽领域、多元化的合作格局，这不仅有力促进了关键资源要素在区域内部实现更加优化配置，也为中国新质生产力培育创造了有利条件。具体而言，“一带一路”共建国家中不乏资源大国，矿产资源储量和种类均十分丰

① 庞瑞芝：《新质生产力的核心产业形态及培育》，载《人民论坛》2023年第21期，第18–21页。

富，且与中国存在较强的互补性。[①] 如印度尼西亚、菲律宾两国的镍矿储备量及出口量分列全球第一、第二位；哈萨克斯坦、塔吉克斯坦等中亚国家拥有丰富的铀、锑、钼金属储量；沙特阿拉伯、阿联酋、科威特等中东国家则是传统油气出口大国。同时，上述资源对中国战略性新兴产业和未来产业的发展均有着重要支撑作用，镍、锑、钼分别在新能源汽车、新材料、航空航天产业中充当关键原材料，铀、石油及天然气等资源则可通过核电、清洁能源等形式，以绿色可持续方式为战略性新兴产业和未来产业供能。伴随共建“一带一路”不断深化及中国新质生产力发展的正向外溢作用发挥，“一带一路”国家一方面将强化自身对华的政策沟通及贸易畅通水平，提升战略性新兴产业和未来产业关键原材料的供给稳定性；另一方面将促进自身资源及矿产精加工水平，带动关键原材料产品供给的质量上升，为新质生产力发展提供更为可靠的助力。

（二）以产业要素互补带动产业转型升级

新质生产力是以新技术与新要素紧密结合的生产力新形态，其形成与发展离不开产业转型升级，[②] 由产业转型升级带动的新型工业化亦能催生新质生产力。[③] 然而，当前中国产业转型升级面临一定阻力。一方面，中国东部沿海地区，尤其是大型城市的产业发展空间目前已趋向饱和，且由于部分产业深度嵌入国际分工链条，对海运运力和物流时效的需求刚性较大，企业向中西部地区有序梯度转移的意愿较弱，导致部分城市用于布局发展战略性新兴产业的土地资源较为有限；另一方面，人口老龄化和工资水平上涨导致由人口红利带来的成本比较优势逐步丧失，制造业整体盈利能力遭到一定削弱，导致部分企业受迫降低研发投入、放缓自主创新发展步伐。上述两方面问题在阻碍产业转型升级的同时，也将对新质生产力的发展带来一定负面影响。

结合上述问题来看，“一带一路”共建国家所具备的丰富劳动力资源和广阔发展空间恰可与中国培育新质生产力、推动产业转型的需求形成互

① 于宏源：《矿产资源安全与“一带一路”矿产资源风险应对》，载《太平洋学报》2018年第5期，第51–62页。

② 石建勋、徐玲：《加快形成新质生产力的重大战略意义及实现路径研究》，载《财经问题研究》2024年第1期，第3–12页。

③ 余东华、马路萌：《新质生产力与新型工业化：理论阐释与互动路径》，载《天津社会科学》2023年第6期，第90–102页。

补，中国可充分利用向“一带一路”产业转移产生的互惠效应，[①] 更好推动新质生产力发展。首先，“一带一路”共建国家劳动力成本优势愈发凸显。在劳动力成本优势推动下，越南、泰国、印尼等东南亚国家已有逐步成为新“世界工厂”的趋势，哈萨克斯坦、塔吉克斯坦等中亚国家的加工制造业规模亦逐渐扩大。其次，部分“一带一路”共建国家产业结构仍偏向低端，具有承接中国产业转移的空间和意愿。当前部分“一带一路”国家仍处于工业化的初级阶段，国内土地、劳动力及自然资源等要素尚未得到充分开发利用，其既具备承接产业转移的充足空间，又具有通过承接产业转移促进本国经济发展、吸收技术外溢的需求。最后，“一带一路”共建国家的制造水平提升可更好满足中国企业上游加工需求。随着近年“一带一路”共建国家对国际分工的融入程度提升，部分国家已积累了充足的制造业劳动力、专业化厂房等资源，整体加工制造水平有所提升。这一方面使得“一带一路”共建国家的生产效率持续提升，可为中国企业上游生产带来更廉价、更高质的供给选择，另一方面也加强了“一带一路”共建国家的产业承载能力，为中国更多相对偏向低端的产业环节提供了转移空间。

（三）以一体化大市场推动供给侧改革优化

从结构承载角度来看，以新兴产业和未来产业等为主导发展形成的现代化产业体系是新质生产力的承载主体，[②] 伴随产业发展的新形态、新路径、新模式形成，中国产业体系发展水平将迅速提升，并带动产能增长和产品种类扩充。值得注意的是，基于中国产业发展历程来看，生产力的迅速发展与长期投资间的期限错配将引发产能过剩问题，[③] 加之现阶段我国有效需求不足问题仍未彻底解决、国内大市场潜能有待充分挖掘，新质生产力的快速发展可能在未来一段时期内带来新的产能过剩问题，2024年政府工作报告在针对新兴产业发展作部署时也提出，要“加强重点行业统筹布局和投资引导，防止产能过剩和低水平重复建设”。可见，新质生产

① 刘友金、周健等：《中国与“一带一路”沿线国家产业转移的互惠共生效应研究》，载《中国工业经济》2023年第2期，第55-73页。

② 黄群慧、盛方富：《新质生产力系统：要素特质、结构承载与功能取向》，载《改革》2024年第2期，第15-24页。

③ 朱安东、张宏博：《科学认识当前我国产能过剩》，载《上海经济研究》2023年第12期，第25-36页。

力培育过程中的产能风险须引起重视。

共建“一带一路”的深化发展能为中国提供充足的海外市场和购买力，从而为新质生产力发展扫清障碍。首先，中国与“一带一路”共建国家贸易基础良好、发展潜力显著。共建“一带一路”十年间，中国与“一带一路”共建国家进出口总额累计达到19.1万亿美元，年均增长明显高于同期中国贸易整体增速和全球贸易增速。[①] 其次，“一带一路”共建国家民众需求层级提升将扩大对中国新产品的需求。近年间，中国对“一带一路”共建国家的援助及投资有效增加了共建国家的就业机会，提升了其国内民众收入水平，实现了区域贫困削减的目标。[②]“一带一路”共建国家民众的收入提升，将带动其需求层级向上攀升，并扩大对除生存必需品外的各类消费品的需求，而提升出口产品种类多元化正是提升产能利用效率的科学对策之一。[③] 最后，中国与“一带一路”共建国家贸易便利化水平的不断提高将促进多边大市场潜力充分释放。中国已成为110多个共建国家的主要贸易伙伴，并与20个共建国家签署了14个自贸协定。可以预期，随着“一带一路”不断发展完善，其区域内多边市场的贸易便利化水平亦将逐步提升，进而助力中国和各共建国家更好挖掘区域市场的贸易潜力，[④] 带动“外循环”高质量发展，为新质生产力培育带来更大动能。

四、发展新质生产力与共建“一带一路”双向赋能的理论框架

（一）发展新质生产力与共建“一带一路”双向赋能的理论基础

发展新质生产力与共建“一带一路”在发展周期和目标导向上高度匹配。新质生产力的形成具有长期性、渐进性特点，前沿技术突破、有效市场规模扩大和产业体系现代化水平提升则是新质生产力形成的重要条

① https://www.gov.cn/zhengce/202310/content_6907994.htm。

② 张原：《中国对“一带一路”援助及投资的减贫效应——“授人以鱼”还是“授人以渔”》，载《财贸经济》2018年第2期，第111-125页。

③ 毛其淋，钟一鸣：《出口多元化如何影响企业产能利用率？——来自中国制造业的微观证据》，载《数量经济技术经济研究》2023年第5期，第113-135页。

④ 孔庆峰、董虹蔚：《“一带一路”国家的贸易便利化水平测算与贸易潜力研究》，载《国际贸易问题》2015年第12期，第158-168页。

件。[①] 这决定了中国在培育新质生产力过程中，需要在长期内持续激发科技创新动能、推动市场扩容及促进产业转型升级。共建“一带一路”高质量发展同样具有长期性、渐进性特点。“一带一路”作为中国推动实现高水平对外开放新格局的重要实践平台，其基本内涵在于推动“一带一路”共建国家实现更加稳定紧密的合作，持续、包容、系统合作是其高质量发展的基本要素，[②] 共建“一带一路”所锚定的“持续合作”正是其长期性、渐进性特征的体现。同时，共建“一带一路”以高标准、可持续、惠民生为目标，以实现更高合作水平、更高投入效益、更高供给质量、更高发展韧性为导向，[③] 而新质生产力以劳动者、劳动资料、劳动对象及其优化组合的跃升为基本内涵，以全要素生产率大幅提升为核心标志，二者均着重强调生产力的现代化、高效化发展，而仅在国别范围上有所不同。

综上所述，发展新质生产力意在促进生产力在新时代下的解放和发展，进而稳步推动高质量发展进程，共建“一带一路”则倡导以合作实现多边范围内的互利共赢发展，其在周期长度、主要目标、关键意义上均存在契合之处，这说明中国新质生产力培育与共建“一带一路”之间具有实现良性互促、双向赋能的理论可能。

（二）发展新质生产力与共建“一带一路”双向赋能的逻辑框架

基于前文分析，本文构建如图1、图2和图3所示的发展新质生产力与共建“一带一路”双向赋能逻辑框架。具体而言，新质生产力培育与共建“一带一路”可通过以下三组良性循环关系实现双向赋能。

1.“技术外溢—资源转化”循环

在这一循环中，新质生产力所带来的技术外溢将带动共建“一带一路”科技化发展，而“一带一路”共建国家的技术水平提升则可从物料供给和技术共研两方面反馈新质生产力发展。

在共建“一带一路”日渐成熟的多边技术合作框架下，中国与共建国家的技术互促发展渠道不断增多。在加快培育新质生产力背景下，可以预期中国在未来一段时期内的科技成果产出水平将有所提升，而在此期间所

① 周文、许凌云：《再论新质生产力：认识误区、形成条件与实现路径》，载《改革》2024年第3期，第26-37页。

② 姜安印、刘博：《高质量共建“一带一路”：特征转变、内涵再构与实现路径》，载《亚太经济》2022年第2期，第104-110页。

③ https://www.gov.cn/zhengce/202310/content_6907994.htm。

取得的成果，将通过日益拓展的“一带一路”科技合作渠道向共建国家广泛外溢，实现对共建“一带一路”的科技化赋能。

“一带一路”共建国家的科技化水平提升，一方面将优化共建国家在关键原材料、零部件等领域内的精深加工和制造水平，强化中国战略性新兴产业和未来产业链条韧性，为科技创新带来稳定环境；另一方面，共建“一带一路”科技化水平提升亦将带动共建国家在科研、教育、创新等领域发展，进而提升共建“一带一路”科技共研能力，对中国技术创新形成反哺，进而赋能新质生产力的培育和发展。

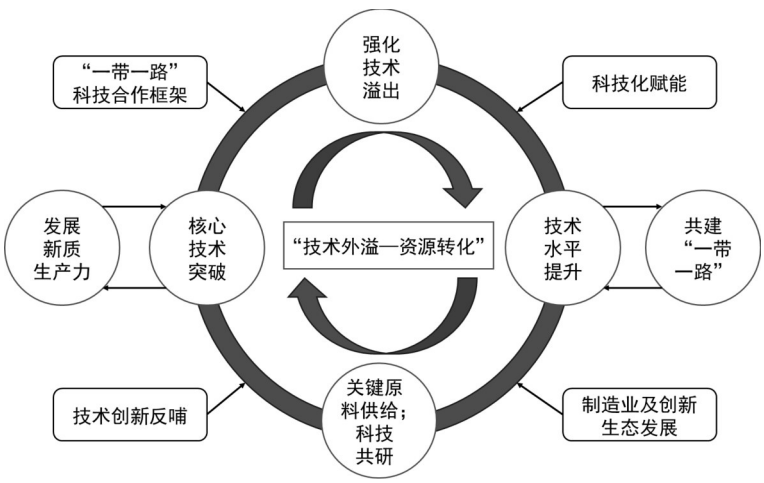

图 1：“技术外溢—资源转化”循环逻辑框架图

2.“产业转移—数字化发展”循环

在这一循环中，新质生产力培育过程中的数字化转型成效将通过产业转移方式向“一带一路”共建国家流动，而共建“一带一路”的数字化发展则将为中国产业的现代化发展提供更大空间。

数字经济所催生的新领域、新模式、新业态是中国培育新质生产力的关键基础，伴随新质生产力的迅速发展，数字经济与各类产业尤其是战略性新兴产业和未来产业的融合程度将不断提高，各类产业数字化转型升级进程亦将加快。在此情况下，战略性新兴产业和未来产业的发展空间需求将持续增加，与当前中国核心产业集聚区的土地资源相对不足之间的矛盾将更为显著。与此同时，部分传统产业或数字化转型相对缓慢的产业则由

于土地租金、劳动力价格上涨而加速向外转移。在产业数字化转型升级加速和相对传统产业向外转移两方面因素影响下，“一带一路”共建国家作为中国产业转移的主要承载，其在吸纳产业资源、带动国家经济发展的同时，也可通过学习中资企业的生产和管理经验，获得部分由产业数字化转型带来的红利，促进自身数字化发展进程。

“一带一路”共建国家的数字化发展进程加快，亦将从几个方面为新质生产力发展提供助力。首先，将推动共建国家形成更高效的生产力，进而带动本国产能优化和优质产品出口增加，使中国产业链条上游供给获得更大的规模和成本优势；其次，将促进共建国家产业体系现代化水平提升，并使其具备承接更多产业环节转移、更高效承接产业转移的能力，进而促进中国产业转移有序发展，助力中国产业体系现代化水平提升；最后，将营造更良好的数字化产业环境，使得中国企业在转移部分产业环节后，仍可通过物联网等技术实现对产业链上下游环节进行管理调节，有助于中国企业的跨境运营效率提升。

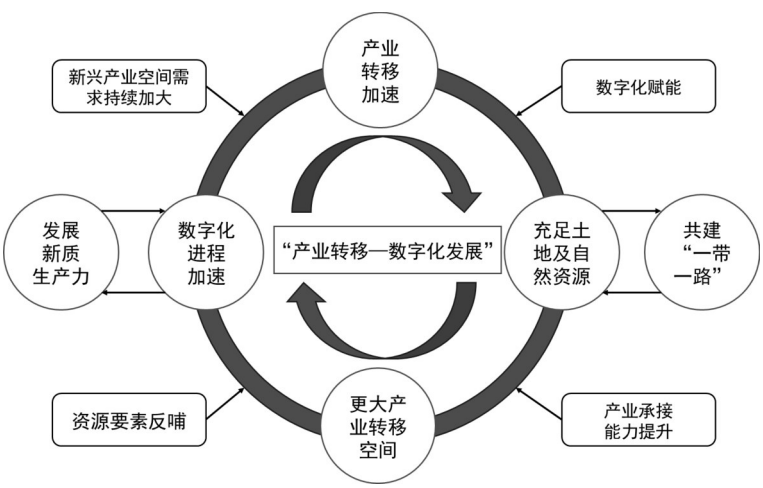

图2：“产业转移—数字化发展”循环逻辑框架图

3.“分工优化—多边市场”循环

在这一循环中，新质生产力发展将以优化生产工具、制造技术出口质量的方式带动“一带一路”共建国家更深入地融入国际分工格局，而“一带一路”共建国家的国际分工参与度和地位提升，亦将推动“一带一路”

多边市场提质扩容，消除新质生产力快速发展带来的产能过剩隐患。

新质生产力的发展使“一带一路”共建国家在生产工具、产业技术、消费品等方面获得更广泛、更优质的进口供给，并以此提升生产制造的现代化水平，应对自身可能存在的生产能力落后、社会必需品缺乏等问题，实现对自身资源禀赋和比较优势的充分挖掘和利用，并依托比较优势积极参与至国际分工当中，获得更高的市场利益和国际分工地位。

“一带一路”共建国家在国际分工中的地位优化，则将推动“一带一路”多边大市场的扩容提质。一方面，国际分工参与度及地位的提升将有效提升共建国家的社会福祉，并提高其国内居民的消费能力，进而使“一带一路”共建国家对中国产业进口需求的规模和种类均出现增加；另一方面，由国际分工格局融入带来的生产专业化将使得“一带一路”共建国家的进口依存度上升，进而提升“一带一路”国家推进区域贸易自由化、市场一体化的意愿水平，最终带动“一带一路”区域市场便利化、自由化程度持续上升，为中国产品“出海”提供更大助力。

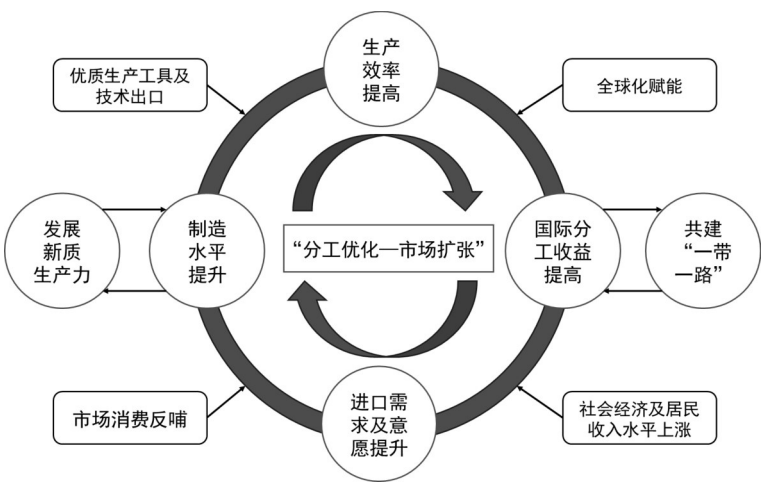

图3：“分工优化—多边市场”循环逻辑框架图

五、发展新质生产力与共建“一带一路”双向赋能的实践进路

当前，全球局势波云诡谲，世界之变、时代之变、历史之变叠加而

来，培育和发展新质生产力与推动共建“一带一路”高质量发展，都将承受较大的外部风险压力。基于此，本文提出以下三方面建议，加快构筑发展新质生产力与共建“一带一路”双向赋能的良性互促格局，助力中国加快形成新质生产力、掌握未来发展主动权。

一是，提升“一带一路”创新能力，打造科技共研共享生态。科技创新既是加快发展新质生产力的核心要素，[①] 也是“一带一路”共建国家全球价值链位势提升的关键抓手。[②] 一方面，中国需持续推进关键核心技术攻坚。以世界科技前沿为目标、以国家重大需求为导向，聚焦芯片技术、精密仪器、工业软件、医疗器械等重大领域进行原创性科研攻关，加快在战略性新兴产业和未来产业的关键环节取得技术突破，不断提升前沿产业链条的自主可控性。另一方面，积极对接联合国2030年可持续发展议程，推动共建“一带一路”创新驱动发展。中国需深化“一带一路”科技合作，带动区域创新能力发展，携手各共建国家共同开展“一带一路”多边科技合作平台建设，放大前沿技术突破在“一带一路”区域内的外溢效应。同时，推进“数字丝绸之路”建设，促进科技同产业、金融深度融合，优化创新环境，推动形成区域协同创新格局，引导技术、人才、信息、资本等要素在“一带一路”共建国家间高效配置、加快集聚，共同营造“一带一路”框架下的优质科创合作生态，推动共建“一带一路”科技水平逐步提升。此外，大力推动“一带一路”科技创新行动计划落地见效，定期开办“一带一路”科技交流大会，不断扩大“一带一路”共建国家联合实验室规模，并积极与各共建国家就人工智能等前沿技术领域的治理展开对话，共促前沿技术健康有序发展。

二是，消弭“一带一路”制度距离，确保双向赋能渠道畅通。良好的体制机制可为新质生产力的形成营造良好的发展环境。首先，中国实施更大范围、更宽领域、更深层次的对外开放，稳步扩大规则、规制、管理、标准等制度型开放，建设更高水平开放型经济新体制，在开放中实现高质量发展，以中国新发展为世界提供新机遇。推动科技创新、要素市场化、人才培育等方面的制度创新，形成有利于新质生产力发展的生产关系，充

① 杨丹辉：《科学把握新质生产力的发展趋势》，载《人民论坛》2023年第21期，第31–33页。

② 张慧智、孙茹峰：《金融发展、科技创新与全球价值链地位》，载《管理学刊》2022年第5期，第1–18页。

分发挥制度对新质生产力发展的保障作用。其次，在制度改革创新基础上，构建多层次政府间政策交流对接机制，推动国内体制机制与“一带一路”各类合作协议形成良好对接；支持各地在因地制宜发展新质生产力的同时，亦开展对接“一带一路”战略行动，不断降低制度复杂性对新质生产力培育与共建“一带一路”之间双向赋能的负面影响。最后，高标准打造“丝路电商”合作先行区，同更多国家商签自由贸易协定、投资保护协定，并全面取消制造业领域外资准入限制措施。主动对照国际高标准经贸规则，深入推进跨境服务贸易和投资高水平开放，扩大数字产品等市场准入，深化国有企业、数字经济、知识产权、政府采购等领域改革，以高效、自由的制度环境推动“一带一路”开放型经济发展。

三是，织密“一带一路”产业合作网络，建设现代化产业分工模式。首先，积极促进数字经济和实体经济融合，带动产业转型升级向高端化、数字化、绿色化、服务化方向发展，同时深度参与全球产业分工和合作，不断增强中国国际经济合作和竞争新优势。同时，增强国内国际两个市场两种资源的联动效应，建设互利共赢、多元平衡、安全高效的全面开放的现代化产业体系，依托中国在全球范围内制造业体系最为完备、规模最大的基础优势，发挥新质生产力培育对“一带一路”国家的外溢效应。其次，紧抓“一带一路”共建国家工业化、城镇化进程加快机遇，深化“一带一路”框架下的产能合作，充分利用各共建国家比较优势，以合作形式突破产业发展瓶颈、实现产业结构升级，共同推动国际分工体系向更高效率、更公平合理的方向发展，与各共建国家携手应对新一轮科技革命和产业变革带来的挑战。最后，构建“一带一路”立体互联互通网络，推进中欧班列高质量发展，参与跨里海国际运输走廊建设，会同各“一带一路”共建国家搭建以铁路、公路直达运输为支撑的亚欧大陆物流新通道。积极推进“丝路海运”港航贸一体化发展，加快陆海新通道、空中丝绸之路建设，为“一带一路”产业分工合作提供更高效、更紧密的交通运输纽带。

编者按：人工智能技术在全球范围内的飞速发展与广泛应用，对全球社会经济、伦理法律、国际关系、国家安全各个领域带来了复杂、深刻而巨大的影响。如何构建有效、公正的人工智能全球治理体系已成为亟待解决的重大课题。为此，本刊特邀国家信息化专家委员会网络空间国际治理研究中心副主任、上海国际问题研究院公共政策与创新研究所副所长鲁传颖研究员，一起组织策划了本期“全球人工智能治理”研究专题，共同探讨人工智能全球治理的理论范式、战略框架和治理体系。

策略与力量：商业主体在人工智能军事化进程中的作用及影响

张璐瑶　鲁传颖

摘　要：人工智能军事化中商业主体的参与是一个重要而特殊的现象。这一新式主体不仅充分参与到人工智能军事领域的研发、应用和部署的全过程，而且对国防能力的提升和国家战略的部署起到重要作用，影响一国人工智能军事化进程的整体走向。对美国、俄罗斯、以色列和法国的案例分析表明，商业主体在军事化结构和进程中的不同作用塑造了各国人工智能军事化节奏和结构的不同特征，并影响国家国防军事能力和部署能力创新的效果。国家需要根据自身的实际情况，调整与商业主体之间的关系，在弥补传统国防工业系统中的短板、激发军事系统活力的同时，引导商业战略与国家安全战略二者和谐互动，共同为负责任的人工智能军事应用做出贡献。

关键词：人工智能军事化；大型科技企业；军工复合体；国防创新

作者简介：张璐瑶，复旦大学国际关系与公共事务学院博士研究生；鲁传颖，上海国际问题研究院公共政策与创新研究所副所长、研究员。

本文系国家社科基金重点项目“网络空间政治安全风险防范与应对能力现代化研究”（23AZD069）的阶段性成果。感谢匿名评审专家和编辑部对本文提出的意见和建议，文责自负。

随着新一轮地缘冲突的爆发，在俄乌、巴以战场上初露锋芒的人工智能军事技术真正走入大众视野，人工智能在战争中的使用问题也脱离了单纯的科幻想象和虚渺的技术恐慌，成为了影响国际安全和亟待全球治理的重要议题。在自动化和智能化武器初步涌现、军事防务领域的人工智能安全事件密集出现、[①] 大国的国防创新战略层出不穷、新型军事合作集团逐步建立的同时，一个更引人注目的新现象逐步显露：有着多样化技术资源和先进技术能力的科技企业开始借由与国家军事部门新建立的密切联系，以一种不同于传统军工企业的新形态参与到人工智能军事化的进程中。[②] 这些商业主体的参与不仅可以影响技术的未来方向，[③] 甚至还能左右国家的战略格局。[④] 因此，有必要对商业主体在人工智能军事化中的角色、作用及其影响进行全面、深入的分析。

本文将从以下几个方面展开论述。首先，介绍人工智能军事化的背景和发展趋势，分析商业主体在这一过程中的特殊作用；其次，梳理人工智能军事化中商业主体的参与脉络，识别其与国家军事战略的互动模式；最后，结合现实分析美国、俄罗斯、以色列和法国四种不同的人工智能军事化模式。这四个国家中商业主体参与程度和范围的不同导致其军事化战略的最终呈现出现了差异。通过对比分析，本文将最终为读者提供人工智能军事化的系统结构和现实图景。

---

① 根据经合组织人工智能事件检测（AIM）数据库，按照“Lethal Autonomous Weapon”“Military Intelligence”“Military Robot”“Defense”“Military”等关键词检索，相关人工智能事件的标准化数量自2023年3月以来开始突出拔高，“OECD AI Incidents Monitor（AIM）”，OECD AI Policy Observatory，https://oecd.ai/en/incidents。

② 参见：孙海泳：《美国人工智能军事化的发展态势、风险与应对路径》，载《国际论坛》2022年第2期，第33-49页；“Why Tech Companies Can No Longer Ignore Their Role in Shaping Politics and Society”，Global Voices Advox，April 19，2023，https://advox.globalvoices.org/2023/04/19/why-tech-companies-can-no-longer-ignore-their-role-in-shaping-politics-and-society/。

③ 参见：M. Scanlon et al.，“16th European Conference on Cyber Warfare and Security（ECCWS 2017）”，August 2017，https://www.proceedings.com/35467.html；François-Xavier Meunier and Renaud Bellais，“Technical Systems and Cross-Sector Knowledge Diffusion：An Illustration with Drones，”*Technology Analysis & Strategic Management*，Vol. 31，No. 4，2019，pp.433-446.

④ 参见：Raquel Jorge-Ricart，“Big Tech Companies and States：Policy or Politics？”Elcano Royal Institute，March 2，2020. https://www.realinstitutoelcano.org/en/blog/big-tech-companies-and-states-policy-or-politics/；黄河、周骁：《超越主权：跨国公司对国际政治经济秩序的影响与重塑》，载《深圳大学学报（人文社会科学版）》，2022年第1期，第107-120页。

一、商业主体在人工智能军事化中的特殊作用

在人工智能军事化的过程中，商业和市场等非传统力量实现了大规模深度参与。数据预测近年来的军事人工智能市场规模将呈指数级增长，从2023年的45.3亿美元增长到2024年的63.8亿美元，复合年均增长率（CAGR）为40.8%。[①] 在如此庞大的市场规模中，微软（Microsoft）、国际商业机器公司（IBM）、帕兰泰尔技术公司（Palantir）等不属于传统军工企业的大型科技企业开始发挥重要而强大的作用。相比于“军工复合体”的传统政企合作模式，这些新兴科技企业在军事领域的作用出现了以下重要调整。

首先，在人工智能领域中，国防部门与新式商业主体在研发合作中的“非对称”关系产生了一定程度的调整。在传统军事领域中，政企双方虽然存在相互依赖的关系，但是在研发环节中，主要由政府的军事部门通过采购合同和国防订单确定和下发任务，传统的国防承包商在很大程度上处于被支配的地位。而在人工智能领域中，政府的军事部门与私营企业的合作方式却能够进一步深入。不仅私营企业中的科技人员可以直接担任国防技术研发中的一些重要职务，而且企业在争取国防订单时有了更大的话语权。[②] 2023年底，美国国家安全局启动了三十年来最大规模的招聘浪潮，预计将招募3000名来自各行各业的技术人员以解决诸如人工智能与国家安全相关的问题。时任国家安全局局长陆军上将保罗·中曾根（Paul M. Nakasone）表示，国家安全局招募和留住顶尖技术人才的能力是满足未来需求的关键。[③] 此外，除了SpaceX和Palantir等大型科技企业在2023年分别拿下了7000万美元和2.5亿美元的巨额国防订单，成为非传统军工企

① “Artificial Intelligence in Modern Warfare Global Market Report 2024”, The Business Research Company, February 2024, https://www.researchandmarkets.com/reports/5896008/artificial-intelligence-in-modern-warfare-global#rela2-5751734.

② 《五角大楼与硅谷企业组建“AI军工复合体”?》,环球时报,2023年12月22日,http://www.news.cn/mil/2023-12/22/c_1212318803.htm。

③ Joseph Clark, “NSA Focuses on Talent as Pace of Technology Quickens”, US Department of Defense News Release, December 8, 2023, https://www.defense.gov/News/News-Stories/Article/Article/3612056/nsa-focuses-on-talent-as-pace-of-technology-quickens/.

业进入美国军事领域中的典型案例以外，[①] 就连中小型的科技企业也有了参与军工领域的能力和机遇。例如，美国陆军通过小型企业创新研究计划（SBIR），拨付了近600万美元的人工智能专项资金给39家小企业。[②] 该计划旨在从非传统小型企业获取人工智能的军事解决方案，涉及自然语言处理、数据分析以及从图像识别到态势感知等方面。本轮获奖公司包括Accrete AI Government、AeroCharge Inc.、ANDRO Computational Solutions LLC、Black Cape Inc.等，这些中小型企业有些甚至是初创的科技企业，凭借其在特定领域的技术能力，给美国的国防创新带来了丰富性与更多可能性。

其次，相比于传统军工企业，人工智能领域中的商业主体的“灵活性”增强。在传统领域中，军工企业的生产模式已经基本脱离了传统的商业赛道，他们需要以争取国防订单为中心安排自身的业务和策略，以获取“军工复合体特殊关系中的经济、政治和社会效益。”[③] 然而在人工智能领域，商业主体参与国防研发时并不会完全脱离市场环境。一方面，这些商业主体可以将其民用领域的技术能力投射到军事领域中，这不仅加快了人工智能军事领域中终端产品的产出速度，还丰富了军事应用的表现方式；另一方面，这些企业还可以充分利用其参与军事项目积累的技术、声誉资源，换取更大的商业利益。为加速人工智能在情报和态势感知上的实战应用，美国国防部发起了Project Maven项目，先后将谷歌、亚马逊和微软等企业拉入项目规划中，利用这些企业在数据标注、场景识别等方面积累的商业经验，为实战装配的效率和质量提供保障。[④] 与此同时，与军方的合作不仅没有打乱这些企业的产品开发节奏，还加速了人工智能领域知识和

① Michael Sheetz，“SpaceX Wins First Pentagon Contract for Starshield，Its Satellite Network for Military Use”，CNBC，September 27，2023，https://www.cnbc.com/2023/09/27/spacex-wins-first-pentagon-contract-for-starshield.html.

② Daniel Smoot，“Army Awards Nearly $6 Million for AI/ML Technologies”，Office of Army Prize Competitions and Army Applied SBIR Program，May 8，2023，https://www.armysbir.army.mil/news/award-6-million-ai-ml-technologies/.

③ 章节根、沈丁立：《军工复合体对美国军控政策的影响》，载《美国研究》2004年第02期，第25-39页。

④ Cheryl Pellerin，“Project Maven to Deploy Computer Algorithms to War Zone by Year’s End”，US Department of Defense News Release，July 21，2017，https://www.defense.gov/News/News-Stories/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/.

技术横向传播的速度，[③] 为这些大型科技企业取得下一波商业产品的创新研发积累了技术能力。

最后，商业主体在人工智能军事领域基本实现了“全流程”参与。这些商业主体可以借由其掌握的技术产品和与军事部门的特殊关系，直接参与到军事行动甚至战略部署中。在传统的军工复合体关系中，企业在按照国防订单交付产品之后，就基本完成了整个合作流程。但是在人工智能领域，原本完全由国家战略和军事部门主导的部署策略却也能被科技企业涉足。在俄乌战场上大显身手、用以态势感知和情报分析的Meta Constellation，就是由Palantir开发并提供给乌克兰军方的。作为一个商业主体的Palantir在此次地缘冲突中却多次表明了其“政治立场”：不仅向乌克兰政府保证其产品是负责、可靠的，而且其首席执行官卡普（Karp）还成为第一位在乌克兰拜访泽连斯基总统并宣布与乌政府达成协议的企业高管。这不仅使得该公司在国际上名声大噪，而且还带来了巨大的商业利润。据报道，Palantir股价在一天内上涨了10%。[①] 这些商业主体在人工智能军事化中的重要地位和作用也成为了加载在当代地缘战略态势天平上的砝码，将会对未来国家之间的战略博弈产生重要作用。

综合以上三点论述，商业主体这种主动性、创造性、全流程的参与方式意味着其扮演了不同于一般意义上的军工复合体在军事化中的新角色。为了更加清晰直观地了解这种新结构是如何形成并发挥作用的，本文接下来将尝试归纳商业主体参与人工智能军事化的能力作用模型。

二、商业主体参与人工智能军事化的过程机制

在常规的国际关系语境下，“军事化”往往指的是将非军事领域的资源投入到军事领域中，并通过沟通和军备常规化使战争行为合法化的过

① François-Xavier Meunier and Renaud Bellais，“Technical Systems and Cross-Sector Knowledge Diffusion：An Illustration with Drones，”*Technology Analysis & Strategic Management*，Vol. 31，No. 4，2019，pp.433-446.

② Jason Novak，“Palantir Gets a 10% Boost on Ukraine Deal，is it Worth Buying the Stock Though？”，Vector Vest，May 26，2023，https://www.vectorvest.com/blog/hot-stocks/palantir-gets-a-10percent-boost-on-ukraine-deal/.

程。[①] 然而除了这个宏观概念以外，具体到人工智能、网络空间等特定领域，“军事化”则应当指将相关的技术资源、研发能力和配套环境开始或有倾向性地投入到军事领域中的过程。以下将从研发、应用、部署三个环节分别对人工智能军事化进程中的行为体、环境、目标、行为事实、进程、联系与反馈进行识别和分析，以此建构一般意义上的商业主体参与人工智能军事化的作用模型（见图1）。

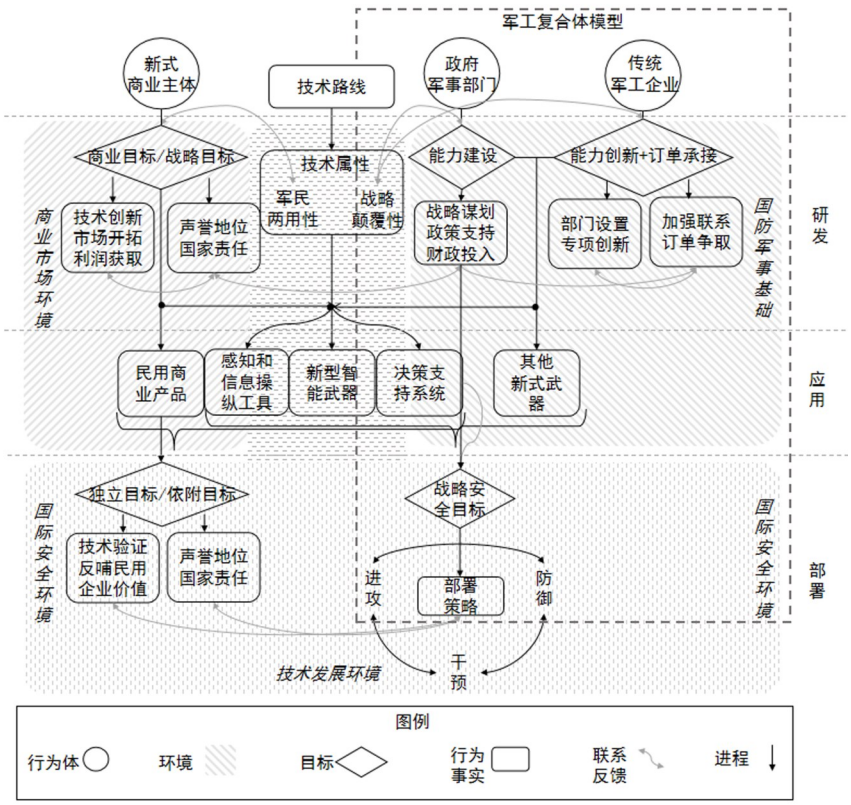

图1：人工智能军事化过程中商业主体作用模型示意图

图片来源：作者自制

从模型中可以看出，对科技企业等新式商业主体在人工智能军事化进程中的作用需要从三个方面进行综合考量：第一，新式商业主体的参与构

① Anna Stavrianakis and Maria Stern，“Militarism and Security：Dialogue，Possibilities and Limits”，*Security Dialogue（Special Issue）*，Vol.49（1-2）：pp.3-18.

成了独特的行为体“结构”，这一结构是整个军事化进程的出发点，商业主体在这一结构中突破了传统军工复合体的一般模式，进一步推动不同国家人工智能领域军事化呈现出不同特征；第二，新式商业主体有效参与了人工智能军事化的完整“进程”，在研发—应用—部署的整个流程中，新式商业主体与军事领域中政府军事部门和传统军工企业等主体建立了较强的相互依赖关系，从而影响国家人工智能军事化领域呈现出不同的走向；第三，新式商业主体的参与推进了整个军事化系统中的决策环节的互动，形成了有效的联系与反馈机制，从而对不同国家军事化的“效果”产生了重要影响。

综合以上三个观测方面，接下来将沿着人工智能军事化进程的脉络，对上述模型进行进一步的解释和说明。

（一）研发环节

人工智能本身的技术属性创设了独特的军事化进程。一方面，人工智能是具有战略意义的军民两用技术，能够从具备一定基础的商业市场环境中吸收经验，推动军事领域与民用领域的融合发展。这在客观上为商业主体参与人工智能军事化进程提供了可能性。另一方面，人工智能是具有强大赋能作用的颠覆性技术，能够推动新型作战概念的发展，从而在战略层面重塑军事力量对比和战争形态。从这个意义上来讲，人工智能技术的发展不仅能够提升传统作战领域的能力，更能够开创新的作战领域，因此引起了国家的战略重视。这在主观上创设了国家推动人工智能军事化发展的必要性。

因此，在整个军事化进程的“研发”环节中，人工智能技术的两大重要属性创设了独特的技术发展环境，并凝聚起了三个重要的行为体。其中，处于国防军事基础环境中的政府军事部门和传统军工企业自然对接了人工智能军事化流程：对于政府军事部门而言，新技术的发展创造了新一轮国防能力发展与建设的必要性。基于这种考虑，2014年奥巴马政府推出了“第三次抵消战略”，旨在凭借军事人工智能领域的高质量优势，弥补与对手在传统军事技术方面的数量上的差距，从而获取整体上的战略优势；[①] 而相应的，2017年俄罗斯总统普京公开宣称人工智能技术所支撑的自主机器人系统将根本性地转变俄罗斯军队的作战模式，并指出采用军用

① 虞卫东：《美国第三次“抵消战略”：意图与影响比较研究》，载《国际关系研究》2015年第03期：第77–87页。

机器人系统是朝正确方向迈出的关键一步。[①] 自此之后，美、俄、英、日等主要大国的人工智能军事战略纷纷出台，开始通过资金投入和政策支持，进一步推动国防能力基于人工智能技术方面的创新发展；而对于传统的军工企业而言，政府国防创新需求也给这些企业的发展规划和技术能力提出了新的要求。由于国防企业的生产投资和发展方向依附于国家战略和国防订单，因此人工智能技术的发展也自然对国防企业提出了转型要求。例如波音、空客、泰雷兹集团等美欧大型国防承包商都在2015年左右在原有产品线上加入了人工智能相关的武器研发，使得其国防订单不降反升。

而处于商业市场环境中的商业主体的参与则更加值得关注，在整个军事化系统的“研发”环节中，商业主体的目标是双向的：一方面，他们希望能够利用人工智能的军民两用属性，通过自身的商业基础促进军事应用的开发，之后再拿军事领域的技术反哺民用领域。在这一过程中，企业能从军事应用的独特需求中获得反馈，推动其技术的进一步创新和优化。军事应用对设备的耐用性、可靠性有更高要求，这促使企业提高产品的质量和性能，从而进一步获取更强的市场竞争力。另一方面，对于一些商业主体来说，参与国家的军事研发还是一种履行社会责任和国家责任的方式。通过提供先进的技术和解决方案，企业可以从竞争激烈的商业赛道中抽身，获得为国家做出贡献的金字招牌。有了这一块金字招牌的背书，企业也能在商业赛道上更高效地挖掘利润和市场声誉。这两个相互加持、相互影响的商业目标和战略目标构成了新式商业主体参与到人工智能军事化研发环节中的基本动力，而这种动力也与国家军事部门的战略意图相向而行，往往能够得到有效对接。在2010年左右，民用人工智能技术在自动化和数据分析等领域实现突破，开始大规模创造价值之时，军事部门也没有忽视这一重要现象，开始与私营企业展开合作。这一时期，美国国防部高级研究计划局（DARPA）与惠普、IBM合作的“SyNAPSE”人工神经网络处理器项目等研发合作与采购活动开始兴起，使商业路线中的经验为军事领域提供先进的技术解决方案和创新思路。

（二）应用环节

商业主体与传统的两大军事主体在研发环节上的相向而行推动了整个

① ［俄罗斯］普京：“战斗机器人可能会显著改变俄罗斯军队”，［俄罗斯］Izvestia，2017年1月26日，https://iz.ru/news/660335。

军事化系统的向前迈进，多样化的终端应用开始出现。从宏观上看，当前主要有三种类别的人工智能军事化的应用形式。

第一类是能够直接服务于战场交火的新型智能武器，具体指那些能够自主执行任务的智能化军事装备，可以进行自主导航、目标识别和攻击决策。这些武器系统集成了先进的人工智能算法，使它们能够在复杂战场环境中高频次、持续性地执行精确打击。例如美国陆军作战能力发展司令部（DEVCOM）的先进瞄准和杀伤力自动化系统（ATLAS）、以色列国防承包商拉斐尔先进防御系统公司开发的GIL-2手持式火箭弹系统以及由法国国防承包商泰雷兹公司开发的RAPIDFire的定点防御武器平台等。[①] 这些系统利用人工智能进行目标识别、追踪和精准打击，显著提高了作战效率和杀伤精度。

第二类是人工智能赋能的决策支持系统，这类系统通过大数据分析、机器学习等技术手段，进行实时的战场态势感知，并基于概率为军事指挥官提供更加准确和迅速的决策依据。它们能够在复杂多变的战场环境中快速分析情报、评估敌我态势、预测敌方行动，并提出计算范围内最优作战方案。其中最著名的案例是美国DARPA在“深绿计划”（Deep Green）的基础上开发的实时对抗情报和决策项目（Real-time Adversarial Intelligence and Decision-making，RAID），这一项目旨在通过预测敌方行动，利用近似博弈论和对欺骗敏感的算法来增强指挥官的决策能力。该项目专注于城市战斗，目的是为战术指挥官实时提供关于敌人位置、意图和战略的估计，以提高作战效率、安全性和效能。[②]

第三类是战场外围的感知和信息操纵工具。这类工具利用智能算法和大数据收集进行系统分析，以提高对战场周边环境的感知能力，并进一步通过操纵信息流，影响敌方的决策过程。这类技术包括但不限于算法认知战、网络影响力行动以及心理战策略的自动化等。通过对信息环境的控制，这类技术应用能够进一步支持军事行动，增强对敌方意图和行动预测的准确性。这些技术应用的呈现方式有的借用商业领域的人工智能工具，

① Zhang Yulong et al.,“Application of Artificial Intelligence in Military: From Projects View”, paper delivered to 6th International Conference on Big Data and Information Analytics (BigDIA), pp.113-116, 2020.

② Michael Ownby and Alexander Kott, “Predicting Enemy's Actions Improves Commander Decision-Making.” *arXiv*, July 22, 2016.

如DeepFake、Snapchat等，通过捏造虚假信息和进行定向传播辅助正面战场的交锋。此外，当前快速发展的生成式人工智能无疑将进一步降低这一造假的门槛，OpenAI新发布的Sora已经可以根据文本自动生成视频内容，虽然当前这些视频内容被OpenAI进行了严格管制，严禁对外发布，但是可以预想到，这些快速发展的生成式人工智能技术能够在未来对战场外围的辅助能力提供强效支持。

在上述三类应用方式中，商业主体的作用不尽相同。由于商业主体并不会完全背离原有的市场和轨道，因此，他们往往更倾向于把已有的人工智能民用技术产品和研发能力投射到军事领域中的特定部位中，而不是像传统军工企业一样独立地开发完全军事化的应用，这使得他们在态势感知和信息操纵等领域的参与更加深入。例如专注于自动驾驶技术的商业企业可能将其技术应用于军事物流和侦察无人车的开发中。这种转换利用了公司在算法优化、传感器集成和数据处理方面的现有专长，使得这些技术能够快速适应军事要求，提高任务执行的效率和安全性。另一方面，拥有强大图像识别技术的企业可能会将这些技术应用于军事侦察和监视系统，提供更快的识别速度和更高的准确率，以支持情报收集和战场意识。

（三）部署环节

在商业主体参与研发应用流程开发出人工智能技术的军事应用之后，如何进行部署则成为了下一个摆在眼前的问题。在这一环节中，行为体进行决策的环境发生了改变，国际安全的客观形势与国家在这个环境中的主观判断成为最根本的因素，因此发挥主要作用的主体是政府军事部门。国家往往会基于战略意图的确定、实现能力的评估以及对战略态势的判断选择合适的部署方案。而在这一过程中，商业主体也会积极参与，抓住实战机遇进行技术路径的验证和自我影响力的发挥。这是人工智能领域军事化进程中的一个新现象——新式商业主体在部署环节中并不总是完全依附国家的军事战略。因此，如何与企业进行有效互动，消减商业主体带来的负反馈效应成了国家必须面对的重要问题。

具体而言，人工智能在军事领域有三种部署形式，国家往往会综合地采取这三种部署形式来组建国防战略。

第一种是进攻型的战略部署，这种类型旨在通过建立相较于对手更胜一筹的技术优势，以主动出击的方式，识别和利用对手防御或战略决策过

程中的漏洞来实现战略目标。赶在潜在威胁发生之前先发制人地实施打击，从而维护自身的安全、利益或最大限度地降低对手有效应对的能力。[①] 在这一类别的战略部署中，人工智能既是力量倍增的重要武器——在提高进攻效率的同时实现跨域作战，从而系统性击溃对手的防御能力，也是在资源独占与质量优化等非对称情况下的决定性力量。智能武器相比于常规武器而言已经展现出了明显的代际差异，再加上决策支持系统和自动化指挥控制系统的协同作用，能够实现更快速、更精确的战术决策和行动执行，这不仅提高了作战效率，也使得战术层面的决策更加灵活多变，以适应不断变化的战场环境。

第二种是防御型的战略部署，这种类型旨在利用人工智能系统的持续监控和分析潜在威胁的能力，对即将发生的攻击提供早期预警，从而可以及时采取对策。[②] 在这一过程中，人工智能系统不仅仅依赖于静态的数据分析，还能进一步利用机器进行深度学习，以适应新的威胁模式和攻击能力。这种动态学习能力使得防御系统能够预测并识别出APT和零日攻击等难以检测的安全威胁，为国防能力和国家安全做出贡献。

第三种是干预型的战略部署，这种类型的战略部署在常规武器时代并不能介入进攻／防御的二分格局，然而在人工智能的加持下，干预的技术手段日益多元、干预的门槛持续降低、干预的成效进一步凸显，使之也成为了一种不容忽视的类型。干预型的战略部署通过人工智能技术，不仅能够在进攻和防御之间找到一条新的路径，还能够在更广泛的领域内发挥作用，如心理战、信息战以及经济战等。这种部署方式利用人工智能系统的数据分析、模式识别和自动决策能力，对敌方的社会、经济、政治和军事系统进行精准的干预和影响。这种方式在俄乌冲突中被以美国为首的西方国家充分利用。在俄乌冲突中，美国和西欧国家通过与乌克兰的情报共享、决策辅助和对俄罗斯的经济制裁和科技脱钩等综合性、跨领域的手段，将原本由俄罗斯占据优势的天平回拨向乌克兰一方，[③] 向世界展示了

① Wilhelm Agrell，“Offensive versus Defensive：Military Strategy and Alternative Defence，”*Journal of Peace Research*，Vol.24，No.1，1987，pp.75-85.

② Lee Hadlington et al.，“The Use of Artificial Intelligence in a Military Context：Development of the Attitudes toward AI in Defense（AAID）Scale”，*Frontiers in Psychology*，Vol.14，2023.

③ 李巍、穆睿彤：《俄乌冲突下的西方对俄经济制裁》，载《现代国际关系》2022年第04期：第1-9页。

干预型战略部署的重要作用。这种部署方式的发展，标志着现代战争形态向更加复杂和多元的方向演进，同时也对国家的安全政策和军事战略提出了新的挑战和要求。

国家想要更加灵活有效地配置这三种部署方式，就必须与另外一个能在部署环节发挥作用的主体——新式商业主体进行有效配合，以消解系统中的负反馈效应。商业主体能在这一领域中产生负反馈效应的原因有两点：一方面，商业主体有着不完全依赖于国家战略的独立目标，如特定的价值观和对商业利益的考量等。美国乔治城大学安全与新兴技术中心（CSET）曾发布报告分析，在俄乌冲突中追随美国政府进行战略部署的亚马逊、苹果、微软等企业可能会由于与中国市场的深度耦合而并不会完全服从于美国对中国的战略决策。[①] 这说明了商业主体在部署环节中决策的独立性与反馈的显著性。另一方面，国际规则的不明确性为这些商业主体参与军事部署环节提供了更多的操作空间，在俄乌、巴以冲突中微软、太空探索公司等商业主体的行为都没有得到相关的赋权或者约束。这使得企业能够轻易且独立地按照自身的价值判断和利益考量干预国家的战略决策。这种行为很可能会反馈到国家的战略部署中，从而增加国家之间战略稳定关系中的不确定性。

为深入探究商业主体在人工智能军事应用领域中的角色及其产生的具体影响，以下将继续通过分析美国、俄罗斯、以色列和法国的具体案例，揭示在不同国家中，商业主体参与的“结构”和“进程”是如何影响国家人工智能军事领域部署的“效果”的。通过对这些案例的考察，我们可以进一步了解商业主体如何在现实中发挥作用，同时进一步把握这些关键国家人工智能军事化的特征和方略。

三、综合外向型：美国商业主体对人工智能军事化的参与

在人工智能军事化进程中，美国的国防工业进一步融合了传统军工企业、新式商业主体以及政府军事部门的力量，共同构成了一个综合而高效的国防技术创新生态系统，实现了商业市场环境与国防工业基础的

① Sam Bresnick et al.，“Which Ties Will Bind?”，Center for Security and Emerging Technology（CSET），February 2024，https://cset.georgetown.edu/publication/which-ties-will-bind/.

融合。据统计，在美国的人工智能军事化的新式应用中，有一半以上是由其新式商业主体开发的，体现了商业主体在美国国防创新体系中的重要作用。

在美国人工智能军事化进程中，商业主体的参与有以下突出特征：

第一，从结构来看，在传统军工行业的人工智能能力开发进度加快的同时，商业主体逐渐深入参与国防项目，专业的新兴人工智能军工企业加速涌现。与美国国防部合作密切的洛克希德·马丁公司（Lockheed Martin Corp.）、波音公司（The Boeing Co.）以及通用动力公司（General Dynamics Corp.）在人工智能领域持续发力：洛克希德·马丁公司设置了先进技术实验室（ATL）并出台了自身的人工智能计划，试图在自适应电子战系统（BLADE）、任务有效性和安全评估系统（MENSA）、智能环境系统（SUSIE）等方面给美军提供能力支撑。[①] 这些传统老牌的军工企业更了解美国军事装配的节奏，对人工智能军事化进程进行了直接有效的跟进，是美国在2010年之前推动人工智能军事化进程的主体。

然而，随着人工智能商业应用场景的丰富，美国国防部越来越注重发掘新式商业主体在人工智能军事化领域中的能力。但是这一过程并不那么顺利，科技企业在参与人工智能军事化时仍然有一定的顾虑和考量。2018年，美国国防部与谷歌签署了著名的Project Maven项目，但当时谷歌基于自身的商业逻辑和道德立场，最终叫停了这一项目。谷歌的诸位员工纷纷表态称，人工智能不应该用以制造战争工具，并推动谷歌退出了联合企业防御基础设施（JEDI）项目。而后续参与Project Maven项目的微软和亚马逊也产生了纠纷和矛盾，这些商业主体在参与军事项目时，对彼此之间的功能和利益协调出现了一系列分歧。然而随着人工智能技术的进一步发展，商业主体不仅看到了参与国防工业的好处，同时也对自身的能力和发展战略进行了调整和优化。2022年，升级版的联合作战云能力（JWCC）项目最终顺利敲定。此时，谷歌、亚马逊、甲骨文和微软四大科技巨头联手斩获了国防部90亿美元的巨额合同，[②] 成为了美国人工智能军事化

① Marcus Roth，“Artificial Intelligence at the Top 5 US Defense Contractors”，January 3，2019，https://emerj.com/ai-sector-overviews/artificial-intelligence-at-the-top-5-us-defense-contractors/.

② “Pentagon splits $9 billion cloud contract among Google，Amazon，Oracle and Microsoft”，Reuters，December 8，2022，https://www.reuters.com/technology/pentagon-awards-9-bln-cloud-contracts-each-google-amazon-oracle-microsoft-2022-12-07/.

中商业主体参与的典型案例。

除此之外，在美国的人工智能军事应用领域，还涌现了一批专业化的新型科技企业。2017年安杜里尔工业公司（Anduril Industries）横空出世，在成立仅仅5年的时间里分别拿下了美国特种作战司令部和美国空军价值10亿美元和800万美元的合同。然而，安杜里尔工业公司这种积极与政府和军方展开合作的行为不符合一般的商业逻辑，因此也被称为“科技界最具争议的初创公司”。[①] 这种现象在美国却不是个例，在最富盛名的十大人工智能新兴军工企业中，美国占据了6席，这些新兴科技企业成为了美国人工智能国防创新的重要商业力量支持。

第二，从进程来看，美国的商业主体实现了对人工智能军事化全过程的参与，并通过加强与传统实体之间的配合，打造了一套成熟、完善的人工智能军事生态体系。如前所述，人工智能军事化进程是一个包含研发、应用和部署的完整系统。在这一系统中，商业主体较容易参与的是前两个步骤，即技术研发和产品输出。借由国防部的采购订单和DARPA的创新项目，不少大型科技企业甚至人工智能领域的初创企业都拿到了数量可观的国防订单，积极参与到了各类人工智能军用技术的开发过程中，实现了军事价值和商业价值的双丰收。2023年，Salesforce、Thomas等企业的市值跃升了29%，为美国的人工智能商业市场注入了动力。

此外，美国的商业主体能进一步参与到美国的战略部署中，给美国实施外向型的人工智能军事战略提供了更多可能性。如图2所示，美国的新兴商业主体在人工智能领域的新兴部署策略——干预性部署中起到了极强的支撑作用。在俄乌冲突中，谷歌进一步扩展其“盾牌计划”，为150多个乌克兰政治组织以及新闻出版机构等提供保护。[②] 亚马逊不仅剔除了俄罗斯的用户，还协助乌克兰将一些机构的网络服务放到亚马逊云端上，以保护其免受攻击。而从总体的战略部署来看，美国的人工智能军事战略部署是外向型的，大多数的精力和重点都被放在了干预和进攻的部署

① Joshua Brustein，“Tech’s Most Controversial Startup Now Makes Drone-Killing Robots”，Bloomberg Businessweek，October 3，2019，https://www.bloomberg.com/news/features/2019-10-03/tech-s-most-controversial-startup-now-makes-attack-drones.

② “Defending Ukraine：Early Lessons from the Cyber War”，Microsoft，June 22，2022，https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE50KOK.

领域。这进一步说明美国在网络领域较为成功的前置防御、前出狩猎等策略也被逐步运用到了人工智能的军事领域，成为美国新型国防策略中的突出风格。

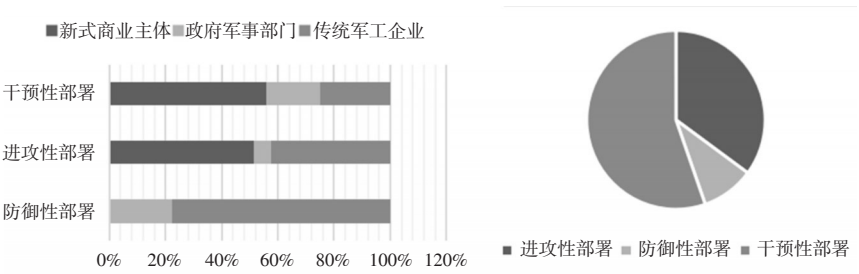

图2：美国的人工智能军事化部署及商业主体的参与结构

数据来源：SIPRI Dataset on autonomy in weapon systems for public release，https://docs.google.com/spreadsheets/d/1M1h2Os7T1UESoSoVpa8ElGXbgvIfjtKD/edit#gid=1786183935

第三，从效果来看，商业主体的参与给美国的人工智能军事化带来巨大收益。首先，在人工智能军事领域，美国再次确立了自身的领跑地位。美国在人工智能军事化应用的产品数量、种类以及军事力量等方面的持续增长，都表明了美国在人工智能军事领域的强大领先地位。[①] 其次，由于商业力量的全方位和深度合作，美国的人工智能军事化潜力巨大。目前，许多硅谷企业都放弃了人工智能非军事应用的承诺，积极参与构筑美国的国防创新系统。OpenAI在2024年1月隐蔽移除了“不允许将其模型用于‘具有高人身伤害风险的活动，包括：武器开发以及军事和战争’”的相关条款，[②] 并且在数日之后公开表示“由于有些国家安全用例与我们的使命相符。OpenAI已经与DARPA开展合作，推动新的网络安全工具的创建”。[③] 越来越多的科技企业深度参与国防工程将为美国在人工智能领域的持续领跑注入强劲动力。

---

① “GlobalFirepower.com Ranks”，GFP，https://www.globalfirepower.com/global-ranks-previous.php.

② “OpenAI Quietly Removes Ban on Military Use of Its AI Tools，”CNBC，January 17，2024，https://www.cnbc.com/2024/01/16/openai-quietly-removes-ban-on-military-use-of-its-ai-tools.html.

③ Jessica Lyons，“Pentagon using ChatGPT？ Oh sure，for cyber-things and veterans，says OpenAI”，the Register，January 16，2024，https://www.theregister.com/2024/01/16/us_military_openai/.

四、路径依赖型：俄罗斯、以色列商业主体对人工智能军事化的参与

近年来，在俄乌、巴以两场新型地缘冲突中崭露头角的智能化军事技术让世界的注意力再度落到了俄罗斯和以色列这两个传统军事强国之上。与美国不同，这两个国家的人工智能军事化路径基本还遵从着传统军事化路径的一般模式，由于商业主体的参与度不够，因此在效率和可持续性上存在一定的短板。

第一，从结构上来看，俄罗斯和以色列的人工智能军事化仍然主要依靠政府军事部门和传统军工企业的智能化转型。然而在这一现象背后，两国的原因却有所差异。对于俄罗斯而言，其人工智能的商业基础较差。根据2023年全球人工智能指数，俄罗斯的人工智能商业评分仅有1.7分，在62个被统计国家中排在第52名。[①] 因此，即便俄罗斯意识到了商业主体在人工智能军事化中的重要作用，但是由于其市场和商业能力的明显短缺，因此也只能采用自上而下的传统军事化方法：2021年，俄罗斯国防部创新发展总局（GUIR）成立了专门的人工智能部门，协同数百个研究所、设计局和测试中心进行人工智能军事应用的研发。到2022年底，GUIR已支持超过500个项目，实现了在指挥、控制、通信和决策、无人驾驶、核武器和高精度武器、防空预警、电子战和天基系统以及网络和影响力行动等多个领域的广泛智能武器开发。这些武器已经在俄乌战场上初露锋芒，例如用以自动化态势感知的RB109-A Bylina、用以自主对空攻击的S-350 Vityaz以及重量级的KUB-LA神风特攻队无人机等。[②] 由此可见，俄罗斯的人工智能军事化进程依赖其强大的军事基础，并且能够在国家和传统军工企业的努力下有效推进。就目前的情况来看，俄罗斯尚未在这一军事改革的浪潮中明显掉队。

而与俄罗斯不同，以色列有着坚实的人工智能商业基础，在同一人工

① Serena Cesareo and Joseph White "The Global AI Index", Tortois, https://www.tortoisemedia.com/intelligence/global-ai/.

② Katarzyna Zysk, "Struggling, Not Crumbling: Russian Defence AI in a Time of War", UK the Royal United Services Institute (RUSI), November 20, 2023, https://rusi.org/explore-our-research/publications/commentary/struggling-not-crumbling-russian-defence-ai-time-war.

智能指数排行榜中，以色列的人工智能商业能力排在第3名（仅次于中美），且拥有Aurora Labs、Beewise等位列世界人工智能100强的大型科技企业。但是以色列目前的人工智能军事化却鲜少与这些商业主体进行合作。其主要原因有以下两点：一是以色列传统的军工企业实现了快速的能力转型，例如以色列最大的军事承包商Elbit在2017年已经开发了具有智能机器视觉的Sky-Striker无人机，并在当年的巴黎航展上引起了世界的关注。此外，以色列航空航天工业公司（IAI）开发的Guardium自主周边巡逻无人机更是早在2008年就付诸部署。这些传统军工主体的快速反应或许是以色列忽视军民融合的原因之一。二是以色列的人工智能军事化发展具有极强外部性。这种外部性不仅表现在部署的外部性，还表现在研发合作的外部性。以色列虽然不和本国的商业主体进行合作，但是却将目标瞄准了谷歌、亚马逊等世界著名的大型科技企业，通过Nimbus项目与这些科技企业签订了价值12亿美元的协议，为以色列军队提供云服务。这使得以色列调用了更加符合自身需要的商业资源，精准推进人工智能军事化进程。

第二，从进程上看，俄罗斯和以色列商业主体的参与有限，整个军事化进程基本依照对传统军事领域的路径依赖设计，并没有体现出较为明显的突破和改进。俄罗斯和以色列两个国家的人工智能军事化进程都在政府军事部门的引导下，由传统军工企业快速推进。然而，具体到部署策略环节，俄罗斯的部署策略较为均衡，兼顾进攻、防御与干预性技术，而以色列的进攻性和外部性更加突出。这也进一步体现在最近的两场地缘冲突之中，以色列对人工智能技术不加节制的使用引发了更加严重的人道主义危机。在加沙地带的军事行动中，以色列军方依靠名为“福音Gospel”的人工智能系统来帮助确定攻击目标，造成了超过3万名巴勒斯坦人员的伤亡。由于没有商业主体协助政府进行道德反思，人工智能的使用权被完全控制在了执行进攻性战略的以色列政府军方手中。

第三，从效果上看，商业力量的不足虽然在当前阶段没有给俄罗斯和以色列两个国家的军事化进程带来突出的问题，但是也显露出了一些隐患。对于俄罗斯而言，这种问题是结构性的。一方面，俄罗斯传统的军事创新结构存在僵化的风险。据统计，俄罗斯当前的国防创新体系容

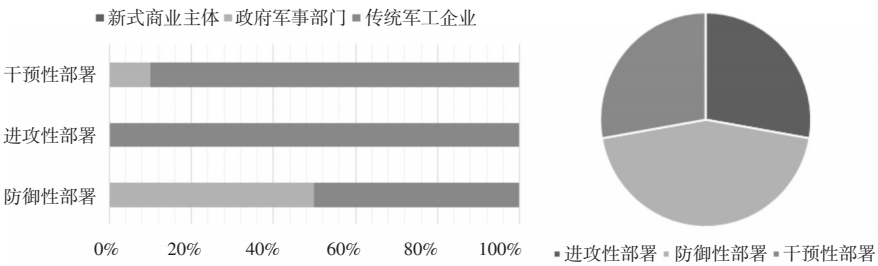

图3：俄罗斯的人工智能军事化部署及商业主体的参与结构

资料来源：作者根据 SIPRI Dataset on autonomy in weapon systems for public release 数据归类整理，https://docs.google.com/spreadsheets/d/1M1h2Os7T1UESoSoVpa8ElGXbgvIfjtKD/edit#gid=1786183935。

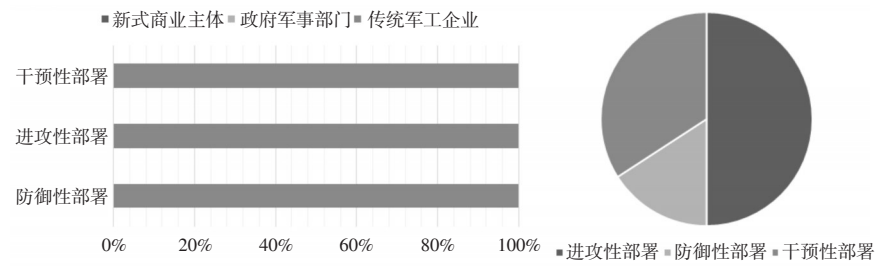

图4：以色列的人工智能军事化部署及商业主体的参与结构

资料来源：作者根据 SIPRI Dataset on autonomy in weapon systems for public release 数据归类整理，https://docs.google.com/spreadsheets/d/1M1h2Os7T1UESoSoVpa8ElGXbgvIfjtKD/edit#gid=1786183935。

纳了近1200个传统的产、学、研实体。[①] 在传统的框架内合作是为了实现资源的集约化和风险的最小化，但是这个庞大的体系却只具备单一中心，对于处于核心位置的俄罗斯国防部创新发展总局（GUIR）而言，如何协调好这些关系是一个具有挑战性的任务。另一方面，在当前俄罗斯面临西方国家全方位封锁的特殊时期，如果还不去思考如何打通商业体系和军事创新体系的壁垒，将有可能彻底丧失人工智能领域技术追赶的机遇期。据统计，俄罗斯的全球创新指数正在逐年下滑，从2016年的第43位下降到2022年的第47位，这暴露了俄罗斯当前军事创新体系中的潜在问题。

而对于以色列而言，这种问题是联系性的。一方面，正如前文提到的，少了商业主体的参与和约束，以色列不加节制的人工智能军事化部署

① ［俄罗斯］俄罗斯联邦国防部："创新活动的结构与实施"，https://mil.ru/mission/innovacia/struct.htm。

正在给全人类奏响噩梦的序曲。另一方面，对于这个战略部署外向型的国家而言，如果一味依赖与其他国家大型科技企业的合作，不去挖掘本国商业主体的国防创新潜力，则可能在国家间战略关系发生转变时受制于人。当前，面对以色列不加节制的人工智能军事化部署，美国国内已有呼吁限制政府和企业与以色列进一步开展安全与防护合作的声音。因此，在平衡军事需求与伦理标准的同时激发国内商业创新，对以色列来说是绕不开的双重挑战。

五、创新干预型：法国商业主体对人工智能军事化的参与

法国同样是在人工智能军事领域表现突出的国家。不同于综合发展引领人工智能军事浪潮的美国，也不同于路径依赖隐含潜在风险的俄罗斯和以色列，法国的人工智能军事化进程创造性发挥人工智能商业主体的优势，塑造了更具特色的国防创新体系。

第一，从结构上来看，法国对商业主体进行了创造性利用，开创了较为实际、灵活的新式军工合作模式。为法国开发了Alister（A Series：A18-M，A-9M，A-27M）、Inspector MK I等新型智能武器的ECA Robotics由于主体企业结构的老旧和竞争力的日渐下降，被引导与iXblue合并为一个新型企业Exail。自成立以来，Exail年营业额达2.5亿欧元，成为前景广阔的新式人工智能企业。而除了引导企业进行合并以提高竞争力以外，法国还将部分企业进行精细化再构建，以充分利用商业主体的独特优势。2012年，M-Tecks Robotics成立，这个小微企业虽然名不见经传，但却是工程和工业机械公司M-Tecks EAC的商业分支机构，专注于为人工智能等难以进入和具有风险的区域提供创新解决方案，也为法国提供了重要的态势感知智能系统Arthron。因此，虽然法国没有世界上首屈一指的人工智能龙头企业，但是创造性地发挥了其商业主体的优势，在法国国防人工智能协调小组（CCIAD）的引导下，法国的人工智能军事创新进程得以快速推进，已然成为了欧洲能力的重要代表。

第二，从进程上看，法国商业主体对人工智能军事化的参与近乎是全流程的，这些商业主体开发的军事化应用支持了法国的干预性部署能力的

发挥，给法国谋取了更多的战略空间。借由法国国防创新部门的CEA/LIST人工智能卓越中心机制，来自学界和商业界的研发人员可以直接参与到人工智能军事化研究和应用等前期环节当中，这种参与方式比起通过国防订单构建的承包关系而言更为高效和直接，为法国在未来战争中的信息化和智能化提供了坚实的技术支持。同时，这些商业实体通过与国家军事部门的紧密合作，能够进一步推动技术发展与国家防御需求之间的同步，从而增强了法国在全球范围内的竞争力和影响力。此外，在应用和部署领域，由于商业主体提供了态势感知技术，法国的军事部署结构得以进一步完善，在人工智能干预性部署领域有了较大的进展（图5）。

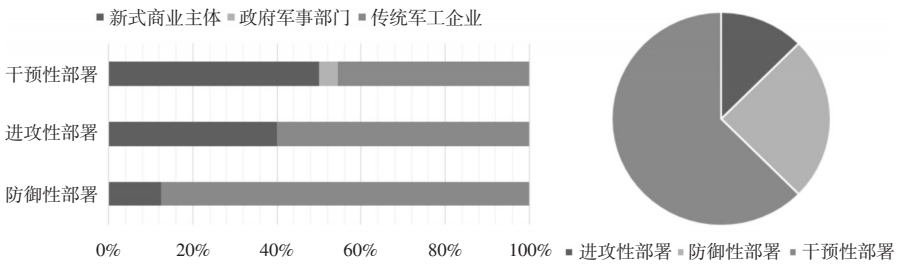

图5：法国的人工智能军事化部署及商业主体的参与结构

资料来源：作者根据SIPRI Dataset on autonomy in weapon systems for public release数据归类整理，https://docs.google.com/spreadsheets/d/1M1h2Os7T1UESoSoVpa8ElGXbgvIfjtKD/edit#gid=1786183935。

第三，从效果上看，法国商业主体的广泛参与不仅加快了人工智能军事化的研发和应用流程，同时反哺了自身的商业市场。例如，前文提到的CEA人工智能卓越中心通过模块化的研究流程设计，不仅提高了军事应用的研发效率，而且实现了军事和商业领域的“双赢”。在过去的5年中，该实验室诞生了20多家人工智能初创企业，极大地提升了法国人工智能的商业力量，形成了军事与民用的有效互补，进而在更为宏观的范围内提升了法国的人工智能竞争力。[①] 这一过程展示了法国如何将国家战略需求与私营部门的创新能力相结合，促进人工智能军事化与商业化的双向流动。不仅增强了国家的防御能力，也推动了科技产业的繁荣和经济增长。通过这样的合作模式，法国在国防安全上保持了前瞻性和先进性的同时，

① “The Challenges of Our Future，”CEA LIST，2020，https://list.cea.fr/app/uploads/2022/04/Activity_report_CEA_List_2019.pdf.

也为商业领域带来了新的增长机会，实现了科技创新与经济发展的良性循环。

六、结语

从现实来看，在主客观因素的共同作用下，各国商业主体在人工智能军事化的实践中，参与程度出现了差异，进而影响了各国的军事部署和国防能力创新的效果。

具有高商业配合度的国家在人工智能军事化进程中表现更为突出，这说明了新式商业主体在人工智能时代军事战略中的重要性。美国、法国等国家不仅具有增强自身智能化军事实力的意图，并且由于科技公司等新式商业主体的充分参与，因而具有了较强的将意图变现的能力。这种军事与商业的高度融合不仅加速了技术的军事化过程，也提高了这些国家在全球安全格局中的能动性和竞争力。而俄罗斯和以色列两个国家则较少依赖商业主体的科技创新，更多选择遵从传统的路径，由国家和传统军工企业主导推进军事化进程，在战略的可持续性、战术的负责任性等方面形成短板，进而给国家安全带来挑战，从长期来看不利于国家智能化军事能力的稳健增长。

商业主体的参与不仅影响国家推动人工智能赋能国防和军事领域的质量和效率，同时也可能进一步推动国家军事化策略的调整。有些国家遵循传统领域对国际安全的承诺，有些国家却有了一定的调整，暴露了其野心，在智能武器的加持下，这可能会给国际安全带来更加严峻的挑战。

欧美人工智能治理模式比较及启示

严少华　杨　昭

摘　要：在国际竞争加剧的今天，人工智能成为了关乎国家安全的关键技术。人工智能治理不仅是各国国内政策的当务之急，也成为全球治理的一个全新领域。欧盟和美国是人工智能技术领域的领先者，也在全球人工智能治理中扮演重要角色。在人工智能治理模式上，欧盟采取的是基于权利和风险的“硬监管”模式，侧重在法律层面搭建全面的人工智能治理框架，以加强监管方面的领导力，发挥其“规范性力量”。美国采取的则是基于市场与技术的“软监管”模式，依赖总统行政令等“软法”进行治理，鼓励“行业自治”和“自我监管”，以保持在技术上的领导力。欧盟和美国在人工智能治理方面的差异具体体现在政策工具性质、治理机构设置以及治理过程中私营部门的参与三个方面。欧美人工智能治理模式的差异可以从政策传统、利益集团影响以及产业生态三个方面得到解释。欧美两种治理模式都是根据各自特殊情况探索出的合适道路，两者之间并不存在根本的冲突。中国具备调和欧美两种治理模式的潜力，找到一条平衡的人工智能治理道路。

关键词：欧盟；美国；中国；人工智能治理

作者简介：严少华，复旦大学国际问题研究院副研究员；
杨昭，复旦大学国际关系与公共事务学院博士生。

人工智能正在掀起新一轮科技革命的浪潮，尤其是ChatGPT等生成式人工智能技术给人类生产和生活带来了革命性的变化。人工智能技术的快速发展在给人类带来前所未有的经济增长与社会进步的同时，其对个人与社会带来的潜在不可控风险也引起了各国监管者的注意。如何在不扼杀创

本文系上海市哲学社会科学一般项目“俄乌冲突下欧洲战略自主走向及其对中欧关系的影响研究”（项目号：2023BGJ003）的阶段性成果。感谢匿名评审专家和编辑部对本文提出的意见和建议，文责自负。

新的情况下对人工智能进行有效的治理成为各国和国际组织需要解决的一个政策难题。

在过去的几年里，世界各国都在加速推出人工智能治理方面的立法，其中既包括全面的立法和专门的立法，也包括自愿遵守的各种指南与标准。根据斯坦福大学的统计，推出含有人工智能内容立法的国家已由2022年的25个增加至2023年的127个。[①] 经济合作发展组织、联合国教科文组织、欧洲委员会（Council of Europe）以及七国集团、二十国集团等也在多边层面出台了人工智能治理的原则与框架。这充分说明人工智能治理不仅是各国国内政策的当务之急，也成为全球治理的一个全新领域。

在各国开展人工智能治理的实践中，逐渐呈现出以欧盟和美国为代表的不同路径和模式。美国采取的是“市场驱动”的模式，欧盟采取的则是“权利驱动”的模式。[②] 欧美人工智能治理模式不仅会重塑其国内市场，也会对中国的人工智能产业及治理产生重要的影响。本文围绕欧盟和美国两种代表性的人工智能治理模式进行比较研究，厘清欧美人工智能治理路径差异并解释差异的原因，以期对中国的人工智能治理提供有益的借鉴。

一、人工智能治理现状与问题的提出

自1956年在达特茅斯被命名伊始，人工智能技术一直经历着以5-10年为周期的循环，每当技术进步出现时会有大批的资金和人力投入，进入属于人工智能的“春天”，但当后续技术发展没有达到既定预期时，便会进入低谷的“寒冬”。[③] 2016年AlphaGo战胜李世石，2022年ChatGPT“横空出世”，技术在短期内的接连进步给人工智能带来了“春天”，也给人类带来了社会进步的无尽遐想。不过，人工智能如同“一个硬币的两

① Nestor Maslej et.al，“Artificial Intelligence Index Report 2023”，AI Index Steering Committee，Inst-itute for Human-Centered AI，Stanford University，Stanford，CA，April 2023.

② Anu Bradford，“The Race to Regulate Artificial Intelligence：Why Europe Has an Edge over Ame-rica and China”，*Foreign Affairs*，June 27，2023.

③ 梅拉妮·米歇尔：《AI3.0》，王飞跃等译，四川科学技术出版社2021年版，第17-18页。

面”，除了机遇还带来了诸多的风险，[①] 也给人工智能治理带来了挑战。

其一，如何区分好的人工智能和坏的人工智能。这种“好与坏”的区分属于伦理范畴。《道德机器》一书早在2012年就讨论了人工智能的伦理问题，指出其是多方面的综合性问题。[②] 2021年联合国教科文组织公布《人工智能伦理问题建议书》，将人权和人的尊严、和平生活、确保多样性和包容性、环境和生态系统蓬勃发展作为四项核心价值观。[③] 不过这一建议书不具备强制性，人工智能伦理还未形成全球性共识。其二，如何处理人工智能并存的机遇与挑战。机遇方面，人工智能可以通过数字赋能促进收入增长、[④] 通过数据共享便利生活、[⑤] 通过在线平台发展远程教育。[⑥] 挑战方面，人工智能可能加大“数字鸿沟”、[⑦] 影响基础教育和高等教育秩序、[⑧] 侵犯个人数据隐私，[⑨] 还可能给国际治理带来治理主体、权力格局和全球安全的不确定性。[⑩] 在治理中发挥技术优势、减轻负面影响，是应对“双刃剑”的核心。其三，如何平衡人工智能的创新和监管，直面

① Breck Dumas，“Google Releases‘AI Opportunity Agenda’for Policymakers”，Fox Business，Nove-mber 14，2023，https://www.foxbusiness.com/technology/google-releases-ai-opportunity-agenda-policymakers.

② 温德尔·瓦拉赫、科林·艾伦：《道德机器：如何让机器人明辨是非》，王小红译，北京大学出版社2017年版，第27页。

③ “Ethics of Artificial Intelligence”，UNESCO，https://www.unesco.org/en/artificial-intelligence/recommendation-ethics.

④ Kate Jones，“AI Governance and Human Rights”，Chatham House，January 10，2023，https://www.chathamhouse.org/2023/01/ai-governance-and-human-rights.

⑤ “Unlocking Value in Manufacturing Through Data Sharing”，World Economic Forum，https://www.weforum.org/projects/data-sharing-for-manufacturing/.

⑥ 苏鹏：《在中国，利用在线工具提升偏远地区的教育水平》，载《信使》2023年第4期，第13-14页。

⑦ Daron Acemoglu et.al，“Artificial Intelligence and Jobs：Evidence from Online Vacancies”，*Journal of Labor Economics*，Vol.40，No.S1，2022，pp.293-340.

⑧ Andre M. Perry and Nicol Turner Lee，“AI Is Coming to Schools，and if We’re Not Careful，so Will Its Biases”，September 26，2019，https://www.brookings.edu/articles/ai-is-coming-to-schools-and-if-were-not-careful-so-will-its-biases/.

⑨ Gary Drenik，“Data Privacy and AI Governance：An Outlook on Tech Industry Trends”，October 12，2023，https://www.forbes.com/sites/garydrenik/2023/10/12/data-privacy-and-ai-governance-an-outlook-on-tech-industry-trends/？sh=768c6f231a55.

⑩ 董汀、黄智尧：《人工智能国际治理与“不确定性”》，载《信息安全与通信保密》2023年第8期，第10-22页。

“科林格里奇困境”。[①] 在国际竞争加剧的今天，人工智能成为了关乎国家安全的关键技术，所以监管与创新的平衡不仅是国内政府和市场关系的协调，还带有国家间竞争的色彩。换言之，人工智能需要创新与监管平衡发展，才能够在为技术进步提供空间的同时保证技术可控。但在国际竞争背景下，先进技术代表的创新和提早监管带来的“先发优势”都可以成为竞争的工具，也让本就是两难的“科林格里奇困境”难上加难。[②]

鉴于欧盟和美国在人工智能治理中的领先地位，针对两者的治理模式对比学界已经有不少研究。其一，总结欧美人工智能治理进展，以人工智能的可信性、安全化、伦理、权利等议题为核心进行总结。[③] 其二，分析对比欧美人工智能治理模式，有学者从“质疑人工智能的权利”或“人工智能可解释性”等角度对比欧盟、美国、英国三者的治理模式，发现美国的政策体系更松散、只有欧盟赋予了公民直接质疑的权利；[④] 还有学者从伦理角度对比欧盟、美国、加拿大的治理模式，发现欧盟最为强调伦理治理。还有研究发现欧美指导文件都关注到了准确性、稳健性、透明度、非歧视性、数据隐私等内容，但监管覆盖范围和监管方式的分歧远大于两者相似之处，相关研究在此基础上分析了跨大西洋人工智能治理的挑战，发现两者存在治理规则错位、数字领域立法交叉影响、人工智能部署性质逐渐复杂等挑战。[⑤]

整体来看，当前研究成果集中于对事实的更新和对政策的总结，主要

① “科林格里奇”困境：一项技术如果因为担心不良后果而过早实施控制，那么技术很可能就难以爆发；反之，如果控制过晚，已经成为整个经济和社会结构的一部分，就可能走向失控，再来解决不良问题就会变得昂贵、困难和耗时，甚至难以或不能改变。David Collingridge，*The Social Control of Technology*（Frances Pinter，1980），pp.13-22。

② Henry A. Kissinger，Eric Schmidt and Daniel Huttenlocher，*The Age of AI and Our Human Future*（John Murray，2021），Introduction.

③ Joachim Roski et.al，“Enhancing Trust in AI Through Industry Self-governance”，*Journal of the A-merican Medical Informatics Association*，Vol.28，No.7，2021，pp.1582-1590；宋黎磊 戴淑婷：《科技安全化与泛安全化：欧盟人工智能战略研究》，载《德国研究》2022年第4期，第47-65页；曹建峰 方龄曼：《欧盟人工智能伦理与治理的路径及启示》，载《人工智能》2019年第4期，第39-47页。

④ Margot E. Kaminski and Jennifer M. Urban，“The Right to Contest AI”，*Columbia Law Review*，Vol.121，No.7，2021，pp.1957-2047；Luca Nannini，Agathe Balayn and Adam Leon Smith，“Explainability in AI Policies：A Critical Review of Communications，Reports，Regulations，and Standards in the EU，US，and UK”，in Proceedings of the 2023 ACM Conference on Fairness，Accountability，and Transparency，A-ssociation for Computing Machinery，New York，USA，pp.1198-1212.

⑤ Alex Engler，“The EU and U.S. Diverge on AI Regulation：A Transatlantic Comparison and Steps to Alignment”，Brookings，April 25，2023.

以此为基础对欧美模式特征进行总结。相比而言，内容描述多，原因分析少。而且，欧美近期都有“里程碑”意义的人工智能政策出台，欧盟《人工智能法案》即将完成立法流程，美国按照拜登政府去年10月出台的总统行政令进行政策部署，基本明确了两者人工智能治理的方向。本文在现阶段对欧美人工智能治理模式进行对比，更有利于厘清双方治理工具、治理机构和治理机制的全貌。同时，本文对治理模式差异的原因进行分析，并总结其对中国人工智能治理的启示。

二、欧美人工智能治理模式比较

欧盟和美国是人工智能技术领域的领先者，也在全球人工智能治理中扮演重要角色。具体到人工智能治理模式上，欧盟和美国却有着诸多方面的差异，主要体现在三个方面。第一，欧盟与美国在人工智能治理政策制定过程中呈现出不同的工具偏好。欧盟重视立法和监管，美国的治理政策则主要出自总统和行政部门，自治偏好明显。第二，两者在治理机构设置上也有不同的格局。欧盟通过《人工智能法案》设计了从欧盟到成员国的专门机构体系；美国机构则有“因事而设”的特点，且集中在总统直属部门。第三，在私营部门对治理过程的参与中，欧盟与美国也有着不同的处理方式。相较于欧盟而言，美国私营部门对治理过程的参与更加活跃与强势。通过比较这三个层面的差异，可以对发达经济体在新兴技术治理领域中的多样性形成更加具体的认识。

（一）政策工具

1. 欧盟

欧盟对人工智能治理的关注由来已久。在欧委会出台《人工智能白皮书》（White Paper on Artificial Intelligence），欧洲议会通过人工智能、机器人和相关技术伦理框架等政策基础上，2021年4月21日欧盟委员会提出《人工智能法案》（Artificial Intelligence Act）提案，由此开始了更具法律约束力、更具监管性质的立法程序。[①] 2022年12月6日欧盟理事会内部达成一

① “Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmoni-sed Rules on Artificial Intelligence（Artificial Intelligence Act）and Amending Certain Union Legislative Acts”，European Commission，April 21，2021，p.2.

致，发布立场文件。2023年6月14日欧洲议会确定一读文件。其后，欧洲议会与欧盟理事会同意开始三方会谈（Trilogue）协调立场。2023年12月8日欧盟的《人工智能法案》第四次三方会谈结束，达成临时协议。2024年2月2日欧盟常驻代表委员会（Coreper）批准综合文本，3月13日欧洲议会批准了综合文本。

《人工智能法案》采用了“平衡和适当的水平监管路径”（Balanced and Proportionate Horizontal Regulatory Approach），呈现横向立法、议题多元、全生态主体参与、基于风险监管的特征。从议题来看，法案涉及金融、医疗、教育、能源、运输、司法等应用领域，透明度、监管沙箱（Regulatory Sandboxes）、数据库管理、后市场监督、信息共享等相关议题也被覆盖。从全生态主体参与来看，提供者、部署者、进口商和分销商全部被列入了监管范围。从风险监管来看，欧盟划分了“不可接受-高-有限-低”的风险框架，主要针对前两种风险类别的人工智能进行监管。从监管力度来看，欧盟规定了罚款范围从750万欧元或全球年营业额的1.5%到3500万欧元或全球年营业额的7%不等。

值得注意的是，当前针对法案内容还存在不少立场争议，给法案执行带来了一定隐患。[①] 争议一是对于风险分类的争议。一方面，对具体风险的界定和监管方式存在争议，禁用人工智能应用清单、高风险人工智能义务尤被关注。[②] 另一方面，欧盟的风险框架划分是否存在合理依据受到质疑，有观点认为不可接受风险、高风险只是欧盟建构的用于全球竞争的政策工具，并非经过严格论证的理性选择。[③] 争议二是对通用目的人工智能（General Purpose Artificial Intelligence）的监管。2023年10月的三方会谈设想针对基础模型进行横向监管和分层

① “Artificial Intelligence Act：Deal on Comprehensive Rules for Trustworthy AI”，European Parliame- nt News，December 9，2023，https://www.europarl.europa.eu/news/en/press- room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai.

② Martin Coulter and Supantha Mukherjee，“EU’s AI Act Faces Delay with Lawmakers Deadlocked After Crunch Meeting”，Reuters，February 16，2023，https://www.reuters.com/technology/eus-ai-act-faces-delay-wit h-lawmakers-deadlocked-after-crunch-meeting-2023-02-16/.

③ Regine Paul，“European Artificial Intelligence‘Trusted Throughout the World’：Risk-based Regulati- onand the Fashioning of aCompetitive Common AI Market”，*Regulation & Governance*，2023，pp.1-18.

监管，[①] 但后续仍破裂。[②] 破裂原因主要是由于法德等欧洲大国反对，[③] 两国分别有Mistral和Aleph Alpha这两家人工智能公司，专注于开发“欧洲版ChatGPT”。达成妥协的终版综合文本中，包含了“通用目的人工智能”，重点关注“具有系统性风险”的通用人工智能，并规定提供者的义务。[④] 争议三是对生物识别的监管。议会和理事会在是否要完全禁止生物识别技术应用方面存在分歧。与理事会不同，议会主张在公共场合也不能够使用生物识别技术，认为这样一来会侵犯公民的隐私权和公共场所的匿名性。在2023年12月的三方会谈中，围绕是否应该允许执法部门使用面部识别或其他类型的生物识别技术这一问题讨论到了凌晨，最终并没有支持议会立场。[⑤]

2. 美国[⑥]

美国政策文件出台的数量、广度、密度也同2018年和2022年两个关键节点重合，2016年在奥巴马政府任期将结束时出台3份文件，以建议性、框架性、侧重“自我监管”为特征，由此人工智能治理在美国联邦政府层面被提上日程。其后，特朗普政府以“保持美国技术领先地位”为基

① Luca Bertuzzi，“AI Act：EU Countries Headed to Tiered Approach on Foundation Models Amid B-roader Compromise”，Euractiv，October 19，2023，https://www.euractiv.com/section/artificial-intelligence/news/ai-act-eu-countries-headed-to-tiered-approach-on-foundation-models-amid-broader-compromise/；Luca Bertuzzi，“EU Policymakers Enter the Last Mile for Artificial Intelligence Rulebook”，Euractiv，October 30，2023，https://www.euractiv.com/section/artificial-intelligence/news/eu-policymakers-enter-the-last-mile-for-artificial-intelligence-rulebook/.

② Luca Bertuzzi，“EU's AI Act Negotiations Hit the Brakes over Foundation Models”，Euractiv，Nov-ember 10，2023，https://www.euractiv.com/section/artificial-intelligence/news/eus-ai-act-negotiations-hit-the-brakes-over-foundation-models/.

③ Andreas Rinke，“Germany，France and Italy Reach Agreement on the Future of AI Regulation in Europe”，Reuters，November 19，2023，https://www.euronews.com/next/2023/11/19/eu-ai-act-germany-france-and-italy-reach-agreement-on-the-future-of-ai-regulation-in-europ.

④ “Proposal for a Regulation of The European Parliament and of the Council Laying Down Harmon-ised Rules on Artificial Intelligence（Artificial Intelligence Act）and Amending Certain Union Legislative Acts- Analysis of the Final Compromise Text with aView to Agreement”，Council of the European Un-ion，January 26，2024.

⑤ Melissa Heikkil?，“Five Things You Need to Know About the EU's New AI Act”，*MIT Technology Review*，December 11，2023，https://www.technologyreview.com/2023/12/11/1084942/five-things-you-need-to-know-about-the-eus-new-ai-act/.

⑥ 鉴于文件性质和数量，本部分重点讨论联邦政府层面的综合性政策文件，专门性政策文件和各州出台的人工智能政策不再讨论。

调出台系列政策文件，延续奥巴马政府特征，同时涉及“政府监管”的内容。受2022年人工智能技术突破影响，拜登政府集中出台大量政策文件。这些文件呈现三个特征。

第一，总统行政令发挥重要作用。特朗普政府和拜登政府都发布了综合性、指导性的总统行政令，以此规定基本治理走向。拜登政府发布的第14110号行政令最为综合，规定了大量的政策计划，包括90天、120天、270天甚至540天不等，相应时间节点会发布计划进度报告，财政部、商务部、国防部、能源部等联邦部门和下属研究院都有各自任务。[①] 另外，从任务性质来看，多为设立治理框架、确定治理基础的建设性任务，有利于保持政策延续性，即使拜登政府不会连任，大概率可以保证下任政府的人工智能政策基调，进行政策“增量”。

第二，重视研发投入。《国家人工智能研究与发展战略计划》（National Artificial Intelligence Research and Development Strategic Plan）在三个政府时期都一脉相承。该文件最早在2016年10月作为《为人工智能的未来做准备》的配套文件提出，2019年6月和2023年5月发布了两版修订版文件。第一版文件提出了人工智能发展计划中的人工智能投资、协同创新、道德伦理、技术标准、人才队伍等七大战略基础；第二版文件新增第八项“扩大公私合作伙伴关系以促进AI发展”；最新一版新增第九项“为人工智能研究中的国际合作建立一个有原则和协调一致的路径”。《2021年美国创新与竞争法》（United States Innovation and Competition Act of 2021），也涉及人工智能领域，授权未来5年投入大约1200亿美元，用于人工智能、高性能计算、量子计算、机器人等关键技术领域的基础及高级研究、商业化以及教育和培训计划，以应对来自中国的竞争压力。

第三，涉及监管但缺乏强制力。从行政令的基调来看，美国人工智能治理的核心是“保持领先地位”，但拜登政府提出“发展前景和风险管控”并重，同特朗普政府相比监管的重要性有所上升。从专门文件来看，《人工智能应用管理指南》（Guidance for Regulation of Artificial Intelligence Applications）和《人工智能风险管理框架》（AI Risk Management Framework）是两份以管理为主的政策文件。前者对人工智能的监管和自治措施都有所

① “Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence”, White House, October 30, 2023.

论述，自治措施占比更多且内容更为详实，监管措施只是提出宏观原则，“政府监管”缺少可操作化解释。后者对人工智能风险和管控进行了详细描述，提出配套措施，但其“自愿实施”的性质决定了缺少强制力，仍然以自治为主。

（二）治理机构

1. 欧盟

欧盟的人工智能治理机构规划全面，以《人工智能法案》为依照，正在欧盟层面和成员国层面建立一系列的决策、执行、监管、咨询机构。

《人工智能法案》出台前已经有一批机构建立。在欧盟层面，欧盟委员会和欧洲议会关注人工智能治理时间较早，分别在2018年设立人工智能高级别专家组（AI HLEG）作为咨询机构和2020年设立数字时代人工智能特别委员会。两个机构具有明确的任务期限，主要职责为设计规划、提供咨询。欧盟也重视抢占人工智能标准制定的“先机”，欧洲电信标准化协会（ESTI）在2023年12月新成立的安全人工智能技术委员会召开了首次会议，计划在人工智能计算平台安全框架方面发布技术规范。

《人工智能法案》为欧洲人工智能治理机构绘制了“蓝图”，也体现了各方对法案的妥协。在欧盟层面，人工智能办公室和欧洲人工智能委员会（EAIB）是核心，两者构成类似欧盟委员会与欧洲理事会的关系。2021年欧委会在提出法案提案时，就设计建立EAIB，各成员国出一名代表组成委员会，欧洲数据保护监督员和人工智能办公室代表列席。EAIB的构想基本延续至2024年2月通过的法案综合文本中，修改之处是主席不再由欧委会任命，而是由成员国选出一人担任。EAIB主要负责协调各国人工智能政策和建议法案实施方式。人工智能办公室最早于2023年6月的欧洲议会谈判立场文件中被提出，2024年1月24日成立，隶属于欧委会的通信网络、内容和技术总局（DG CNECT），预算由“数字欧洲计划”（Digital Europe Programme）提供，但并不如议会所设想能够“具有独立法人资格”。[①] 办公室是欧委会内部主要负责法案实施和执法的部

① “Amendments Adopted by the European Parliament on 14 June 2023 on the Proposal for a Regul−ation of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial In−telligence（Artificial Intelligence Act）and Amending Certain Union Legislative Acts”, European Parliament, June 14, 2023, p.9.

门，需要为EAIB提供秘书处。[①] EAIB和办公室共同构成欧盟人工智能机构的核心，法案规定围绕两者需设立咨询论坛和独立专家科学小组。[②] 在成员国层面，《人工智能法案》要求各成员国应指定国家主管机构和单一联络点（Single Point of Contact），国家主管机构应包括通知部门（Notifying Authority）和市场监管部门，对应委员会的两个子机构，单一联络点应指定一个市场监管部门担任。在国家主管机构成立后，应每两年向欧委会提交人力及财力情况报告，欧委会转交人工智能委员会并讨论相关建议。

2. 美国

美国的人工智能治理机构建立过程具有明显阶段性，存在两个突出节点：一是第117届国会任期内通过了《国家人工智能倡议法案》（National Artificial Intelligence Initiative Act of 2020），规定建立一批人工智能治理机构；二是2022年生成式人智能技术取得突破，影响了一批机构建立。当前机构呈现两个集中，一是机构集中隶属于行政部门，二是机构性质集中为研究、咨询、协调机构。

在机构关系中，绝大多数机构为总统直属机构或隶属于特定联邦部门，立法机构和司法机构中新设立的机构少。总统直属机构包括2018年6月建立的人工智能专责委员会（SCAI）和2021年1月建立的白宫国家人工智能计划办公室（NAIIO）。这两个机构的“升级”体现人工智能在美国国家治理议程中地位的上升。在组织结构方面，NAIIO级别更高，隶属于白宫科学和技术办公室，由副总统领导；SCAI则隶属于国家科学与技术委员会（NSCT）。在职能方面，NAIIO处于“整个美国创新系统中的国家人工智能研究和政策的中心枢纽”，会参与联邦政府未来多年的人工智能工作；SCAI主要负责为白宫提供建议，如果2025年无更新必要则不再继续运行。联邦部门机构中，并未建立专门的人工智能部门或监管机构，但有不少联邦部门在职权范围内发布了相关人工智能政策，尤以国防部和

① Osman Gazi Güçlütürk and Bahadir Vural，“European Artificial Intelligence Office Established”，H-olisticAI，January 26，2024，https://www.holisticai.com/news/european-artificial-intelligence-office.

② 咨询论坛需要兼顾工业界、初创企业、小微企业、公民社会和学术界利益，编写年度报告或按要求提供建议；专家小组需要由欧委会和人工智能委员会共同决定人选，专注风险分类建议、通用人工智能评估等工作。

商务部突出。[①]

从机构性质看，咨询、协调、研究机构多，执行机构少，且尚无监管机构。由于美国人工智能治理仍处于起步阶段，需要大量的咨询机构给出政策建议，因而有一批任务性、阶段性的咨询机构。国家人工智能安全委员会（NSCAI）是根据《2019 财年国防授权法案》（National Defense Authorization Act for FY2019）设立的独立咨询机构，可以直接向总统和国会提出政策建议，2021 年 3 月发布最终报告，任务结束。国家人工智能倡议工作组（NAIRR Task Force）也是阶段性咨询机构，运作时间从 2021 年 6 月开始到 2023 年 4 月完成工作。从规划设计看，美国人工智能治理是“跨部门治理”，相比于“政府监管”更倾向于“行业自治”，[②] 所以目前是协调机构发挥主要作用而非监管机构，除 NAIIO 这一枢纽机构外，2018 年就建立了隶属于 NSTC 的网络与信息技术研发项目（NITRD）和人工智能研发跨部门工作组（AI R&D IWG）。从技术发展看，美国鼓励人工智能技术创新，[③] 研究机构在其中占比很大，绝大多数研究机构隶属于国家人工智能研究院（NAIRI），该研究院隶属于美国国家科学基金会（NSF）。商务部的国家标准与技术研究院也有人工智能研究机构建立，如美国人工智能安全研究联盟（AISIC）。

① 国防部的首席数字和人工智能办公室（CDAO）成立于 2022 年 2 月将国防部的联合人工智能中心（JAIC）、国防数字服务（DDS）、首席数据官和企业平台 Advana 整合为了一个机构。商务部新成立的人工智能治理机构数量最多，有服务于部门和联邦之分。服务于部门，商务部成立了人工智能卓越中心，隶属于商务部的美国专利商标局（USPTO）和美国国家标准与技术研究院（NIST）也成立了新的机构；服务于联邦，美国国家人工智能咨询委员会（NAIAC）在 2021 年 9 月设立于商务部，负责就人工智能相关主题和《国家人工智能倡议法案》相关事宜向总统和 NAIIO 提供建议。

② 有关行业自治的内容可参见：Joachim Roski et.al，“Enhancing Trust in AI Through Industry Self-g-overnance”，*Journal of the American Medical Informatics Association*，Vol.28，Issue.7，2021，pp.1582-1590；Patricia Gomes Rêgo de Almeida，Carlos Denner dos Santos and Josivania Silva Farias，“Artificial Intelli-gence Regulation：A Framework for Governance”，*Ethics Information Technology*，Vol.23，2021，pp.505-525；Pedro Robles and Daniel J.Mallinson，“Catching up with AI：Pushing Toward a Cohesive Governance Framework”，*Politics & Policy*，Vol.51，Issue.3，2023，pp.355-372。

③ 有关创新政策的内容可见：LucaNannini，Agathe Balayn and Adam Leon Smith，“Explainability in AI Policies：A Critical Review of Communications，Reports，Regulations，and Standards in the EU，US，and UK”，*Proceedings of the 2023 ACM Conference on Fairness，Accountability，and Transparency*，2023，pp.1198-1212。

（三）私营部门参与

私营部门在欧盟和美国都参与了人工智能治理，但参与路径有一定差别：欧盟是“政策先行”，在欧盟层面设定统一的人工智能治理框架后，成员国、私营部门和民间社会参与其中，欧盟与私营部门的联系以“人工智能联盟”为代表的官方公私伙伴关系平台为主；美国是“技术先行”，科技公司可以直接参与美国的人工智能新机构建设和政策设计中。

1. 欧盟

私营部门在欧盟人工智能治理中进入政策决策、引导政策走向的作用较弱，更主要是作为治理的参与者、利用者和遵守者。换言之，私营部门不容易表达“希望欧盟怎么做”，更多地是按照欧盟的政策构想进行调整和融入。

在人工智能治理前期，私营部门主要通过AI HLEG参与人工智能政策设计。AI HLEG的52名成员就不乏来自科技公司、研究机构、高校和民间组织的成员。AI HLEG最终出台了四份文件，[①] 对人工智能伦理、应用和后续政策设计都奠定了基础。在试点和实验阶段，AI HLEG会联系私营部门取得意见和建议。但在任务完成之后，私营部门便缺少了参与和影响人工智能政策设计的官方路径，此后公私伙伴关系平台成为了主要的互动模式。

欧盟建立了人工智能联盟（AI Alliance）、人工智能数据中心（DIH4AI）、“以人类为中心的人工智能倡议国际推广”等公私伙伴关系平台。人工智能联盟和人工智能数据中心是专注于欧盟内部的治理平台，两者的构想发端于2018年4月的《人工智能合作宣言》（Declaration of Cooperation on Artificial Intelligence）。联盟在2018年6月召开了第一次大会，旨在建立关于人工智能的开放式政策对话，希望容纳立法者和公民、学者和从业者、公共当局、民间社会、商业和消费者组织等主体。数据中心隶属于“欧洲数字创新中心网络”（EDIH-AICS），为了确保人工智能合乎伦理地发展，欧委会在2019年秋季之前启动了该项目，与成员国和利益相关者一起，开始讨论制定数据共享模型。除了欧盟域内的私营部门，欧盟还注重提升

① 四份文件包括：《可信人工智能道德准则》（*Ethics Guidelines for Trustworthy*）、《可信人工智能的政策和投资建议》（*Policy and Investment Recommendations for Trustworthy AI*）、《可信人工智能最终评估清单》（*Assessment List for Trustworthy Artificial Intelligence*）、《医疗保健制造部门和物联网部门的应用》（*Sectoral Considerations on the Policy and Investment Recommendations*）。

全球影响力，“以人类为中心的人工智能倡议国际推广”是主要媒介。这一倡议由欧盟委员会外交政策工具服务处（FPI）和通信网络、内容和技术总司（DG CONNECT）与欧洲对外行动服务署（EEAS）合作推出，是一个大型外交政策工具项目，旨在与国际社会接洽，在监管和伦理问题上成为合作伙伴，促进可信赖人工智能的负责任发展。

2. 美国

私营部门在美国人工智能治理中有一定的议程设置能力，在信息沟通、机构设置、治理举措等方面都可见私营部门的影响，特别是2022年以来，私营部门参与极为频繁。

见面会、听证会、论坛等是联邦政府、国会和私营部门沟通的主要渠道。拜登政府在2023年5月召开了“负责任的人工智能创新见面会”，OpenAI、Anthropic、微软、谷歌等行业“翘楚”的首席执行官参加。[①] 此次见面会规格颇高，总统、副总统、白宫科学与技术办公室主任、总统国家安全事务助理等人都出席了见面会，并呼吁与会公司为行业做出榜样，让技术发展遵循国家治理政策。国会两院同样多次召开听证会同企业沟通。除了传统听证会外，参议院多数党领袖舒默还举办了人工智能洞察论坛（AI Insight Forum）。根据其在6月宣布的人工智能技术的“安全创新框架”（SAFE Innovation Framework），要以“全员参与”的方式应对人工智能发展。[②] 根据Tech Policy统计，在164名参与者中，有52名来自科技行业。[③]

在新机构设置中有诸多咨询私营部门意见和建议、建立公私伙伴关系平台的设计。在任务型机构中，私营部门参与其中并发挥作用，为未来治

① “FACT SHEET：Biden-Harris Administration Announces New Actions to Promote Responsible AI I-nnovation that Protects Americans’ Rights and Safety”, White House, May 4, 2023, https://www.whitehouse.gov/ostp/news-updates/2023/05/04/fact-sheet-biden-harris-administration-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety/.

② “Statements from the Ninth Bipartisan Senate Forum on Artificial Intelligence”, Senator Chuck Sc-humer, December 6, 2023, https://www.schumer.senate.gov/newsroom/press-releases/statements-from-the-ninth-bipartisan-senate-forum-on-artificial-intelligence. 九次听证会的主题包括：“人工智能创新”,“版权与知识产权”,“用例和风险管理”,“劳动力”,“国家安全”,“防范世界末日情景”,“人工智能在我们的社交世界中的作用”,“透明度、可解释性和一致性”,“隐私和责任”。

③ Gabby Miller, “US Senate AI ‘Insight Forum’ Tracker”, *Tech Policy Press*, December 9, 2023, https://www.techpolicy.press/us-senate-ai-insight-forum-tracker/.

理进行规划或落实路线图。NSCAI的成员有来自私营部门和研究院的成员。[①] NAIRR工作组规定，政府、学术界和私人组织拥有平等的代表权。在咨询机构中，隶属于NSTC的系列机构规定私营部门能够提供建议；NAIAC的组织结构中有私营伙伴。在研究机构中，私营部门大量参与。最典型的如NAIRI，由NSF牵头，与西蒙斯基金会、NIST、国防部、第一资本金融公司和英特尔公司等合作，目前有25个研究机构，连接美国和全球500多个资助方和合作方。此外，在协调机构和行政机构中，也有协调私营部门的职能或私营部门参与的组织结构。

表1：欧美人工智能治理模式特征对比

	政策工具	治理机构	私营部门参与
欧盟	重视立法和伦理，以《人工智能法案》为核心	欧盟—成员国双层架构，设立监管部门	政策先行，私营部门向欧盟政策靠拢
美国	重视行业自治和产业政策，以总统行政命令为核心	机构集中隶属于行政部门，机构性质集中为研究、咨询、协调机构	技术先行，私营部门有一定议程设置能力

资料来源：笔者自制

三、欧美人工智能治理模式差异的影响因素

治理是公共部门和私营部门的共同参与，通过双方互动形成相对固定的治理模式。政策传统体现公共部门的决策，是治理模式的纵向延续。政策影响力体现私营部门的影响，是治理模式的横向变动。除了这两个影响治理模式的共性因素，产业生态是技术治理领域的个性因素，体现公私部门的相互作用。

（一）政策传统影响

欧盟和美国存在两种治理模式之争，反映在价值逻辑层面分别对应“权利驱动路径”（Rights-driven Approach）和“市场驱动路径”（Market-driven Approach）之争。[②] 两种不同的价值逻辑在具体政策中表现为“监

① “NSCAI Staff”, The National Security Commission on Artificial Intelligence, https://cyber-cemetery.unt.edu/nscai/20211005230904/https://www.nscai.gov/about/nscai-staff/.

② Anu Bradford, “The Race to Regulate Artificial Intelligence: Why Europe Has an Edge over Amer-ica and China”, *Foreign Affairs*, June 27, 2023.

管—自治”之争和“技术—规范”之争。这种对新兴技术的治理路径差异从数字技术延续到了人工智能，体现了两种不同的“社会技术想象”。[①]

“监管—自治”之争体现为对新兴技术应该由“政府监管”还是“行业自治”。欧盟对新兴技术监管一贯严格，监管也是推进欧洲科技一体化的一种手段。[②]欧盟通过战略规划、产业政策、法律框架以及监管机制等制度构建，首先建立了治理框架，以此为导向推进成员国政府政策的协调以及科技公司、民间社会的参与。[③]在欧盟政策语境中，人工智能治理属于数字治理的细分领域，体现了治理路径的“惯性”。[④]首先是立法延续性，欧盟在数字领域的现有法规会涉及人工智能的部分，《人工智能法案》也同《数字市场法》（Digital Market Act）、《数字服务法》（Digital Service Act）等类似，偏向一个法律全面监管的“横向”模式。[⑤]其次是伦理延续性，欧盟重视数据伦理，人工智能治理中有不少关于伦理的内容，属于数据治理细分领域中伦理设计最完备的领域，《欧盟机器人民事法律规则》（Civil Law Rules in Robotics）、《人工智能责任指令》（AI Liability Directive）、《可信人工智能伦理指南》（Ethics Guidelines for Trustworthy）以及助推《人工智能法案》实施的《人工智能协调计划》（Coordinated Plan on Artificial Intelligence）都包含伦理的内容。

美国在治理举措中极为重视“行业自治”，但实质上是产业政策引导的自由市场。[⑥]在三届政府中，《人工智能研发国家战略计划》是唯一一个延续了三届政府的人工智能政策，塑造了美国对人工智能的“底色”，

① Rob Guay and Kean Birch, “A Comparative Analysis of Data Governance: Socio-technical Imagin-aries of Digital Personal Data in the USA and EU (2008-2016)”, *Big Data & Society*, Vol.9, No.2, pp.1-13.

② 肖红军、张丽丽、阳镇：《欧盟数字科技伦理监管：进展及启示》，载《改革》2023年第7期，第73-89页；李研：《“框架化”——观察欧盟科技政策的一个重要视角》，载《科学学研究》2021年第9期，第1604-1612页。

③ 肖红军、张丽丽、阳镇：《欧盟数字科技伦理监管：进展及启示》，第73-89页。

④ 同上。

⑤ Matt O’Shaughnessy and Matt Sheehan, “Lessons from the World’s Two Experiments in AI Gover- nance”, Carnegie, February 14, 2023, https://carnegieendowment.org/2023/02/14/lessons-from-world-s-two-experiments-in-ai-governance-pub-89035.

⑥ Sujai Shivakumar, “Priming the Innovation System: A New Age of U.S. Industrial Policy”, Center for Strategic & International Studies, September 29, 2023, https://www.csis.org/analysis/priming-innovation-system.

即重视产业政策，强调技术发展。后续尽管有特朗普政府出台的《人工智能应用监管指南》，以及拜登政府对技术风险的讨论有所增多，但对技术自由放任的基本态度没有改变。国会方面同样存在“路径依赖”。奥巴马政府和特朗普政府时期主要依靠“软法”进行治理，为数不多已通过的法律仍是持产业政策立场，专注于技术发展而非监管和风险管理，[①]这一路径在第118届国会仍然持续，如出台《芯片与科学法案》（CHIPS and Science Act）、《通胀削减法案》（Inflation Reduction Act）等法案。而且，奥巴马同拜登政府和国会议员均有接触，对美国当下的人工智能治理仍产生着影响：奥巴马和舒默等议员曾有会面和讨论如何对人工智能进行最佳监管，拜登和奥巴马也有定期会面；其也参与了拜登第14110号行政令的制定，并发挥了沟通行业领袖、学术界、民间社会和政府之间的作用。[②]

“技术—规范”之争体现为行为体更倾向于追求技术权力还是规范权力。欧盟一直是“规范性权力”，是国际规范和国际标准制定领域的“佼佼者”，也通过“布鲁塞尔效应”（Brussels Effect）利用欧盟市场规模赋予欧盟规范以国际应用和单边监管的优势。[③]随着国际技术竞争的加剧，欧盟开始强调“技术主权”（Technological Sovereignty）。“技术主权”最先源于对数字技术的治理，后扩展至广义的技术领域，同样依赖和服务于欧盟的“规范性权力”，体现为制定和影响技术标准和管理规则的能力、独立开发和维持关键技术的能力以及技术治理意识形态的影响力。[④]在欧盟权力传统和“技术主权”的影响下，欧盟人工智能治理也希望能够发挥“规范性权力”，[⑤]并主要通过法律和标准两种路径实现。通过制定法律的

① Adam D.Thierer，“U.S. Artificial Intelligence Governance in the Obama-Trump Years”，*IEEE Trans-actions on Technology and Society*，Vol.2，No.4，2021，pp.175-182.

② Richard Lwler，“Barack Obama on AI，Free Speech，and the Future of the Internet”，The Verge，November 8，2023，https://www.theverge.com/2023/11/8/23952224/barack-obama-on-ai-free-speech-and-the-future-of-the-internet.

③ Anu Bradford，“The Brussls Effect”，*Northwestern University Law Review*，Vol.107，No.1，2012，pp.1-68.

④ 蔡翠红、张若扬：《“技术主权”和“数字主权”话语下的欧盟数字化转型战略》，载《国际政治研究》2022年第1期，第9-36页。

⑤ 卓华、王明进：《技术地缘政治驱动的欧盟“开放性战略自主”科技政策》，载《国际展望》2022年第4期，第39-61页。

路径，《人工智能法案》是世界首个人工智能监管立法，欧盟希望继续利用“布鲁塞尔效应”抢占对全球人工智能企业的监管先机，甚至引领“第四波法律全球化”。[①] 通过制定标准的路径，欧盟试图强调技术自主和欧盟价值观，欧洲电信标准协会（ETSI）最近的“排外性”标准制定就是例子。[②]

美国对人工智能的监管则面临“困境”，[③] 既要对华竞争技术主导权，又要对欧竞争规则主导权。[④] 一方面，美国主要依靠市场作用完成对高新技术的融资、研发。这一模式着眼于全球技术竞争，如果立即加大对人工智能的限制可能会使科技企业束手束脚。同时，美国尤其关注中国的技术发展，担心加强监管会落后于中国。另一方面，美国国内不断有需要抢占人工智能监管和技术标准先机，否则未来美欧间跨国开发和合作及美国公司进入欧洲市场都必须遵从欧盟标准的声音。OpenAI为代表的科技公司则掀起了“游说风暴”，寻求减少对人工智能和大模型开发的监管。[⑤] 在美欧贸易与技术委员会（TTC）历次部长级会议中，人工智能也是双方协调的重点但进展有限，主要是在概念上进行讨论，体现了美欧的矛盾。而且，美国借TTC这一平台似乎对欧盟的ETSI“排外性”有隐晦批评：“正如2022年的一份报告中所解释的那样，欧洲关于与美国在技术标准方面合作的声明与迄今为止的行动不符，这些行动排除了多年来在欧洲标准制定中发挥建设性作用的外国公司的专家。”[⑥]

---

① 余成峰：《第四波法律全球化?》，载《读书》2023年第5期，第140-148页。

② ETSI在全球拥有900多个成员组织，来自5大洲的60多个国家，其董事会包括众多总部不在欧洲的跨国公司，例如三星、索尼、IBM、AT&T、高通、思科和苹果。以往在制定标准时，董事会的企业无论其总部是否在欧洲都可以参与决策。

③ Natasha Lomas, “EU and US Lawmakers Move to Draft AI Code of Conduct Fast”, Tech Crunch, June 1, 2023, https://techcrunch.com/2023/05/31/ai-code-of-conduct-us-eu-ttc/?guccounter=1.

④ Anu Bradford, *Digital Empire: The Global Battle to Regulate Technology* (Oxford University Press, 2023), pp.169-250.

⑤ Natasha Lomas, “OpenAI's Altman and Other AI Giants Back Warning of Advanced AI as ‘Extin-ction’ Risk”, TechCrunch, May 30, 2023, https://techcrunch.com/2023/05/30/ai-extiction-risk-statement/.

⑥ “TTC Joint Roadmap on Evaluation and Measurement Tools for Trustworthy AI and Risk Manage-ment”, National Institute of Standards and Technology, December 1, 2022, p.3, https://www.nist.gov/system/files/documents/2022/12/04/Joint_TTC_Roadmap_Dec2022_Final.pdf.

（二）利益集团影响

代表人工智能企业的利益集团的政策影响力同样是导致欧盟与美国在该领域的治理模式差异的重要原因。由于主要欧盟成员国与美国均属于竞争性民主制度，政党构成了政治生活中的核心要素，且欧盟领导层和主要机构也由定期的竞选程序决定，因此这构成了包括企业在内的众多利益集团发挥其正式或非正式的政策影响的制度基础。这里的政策影响力是指，企业通过游说、献金、选票等形式影响政党、候选人和公共机构政策偏好的能力。[①] 由于欧盟与美国的政策出台和政治竞争过程存在私营部门发挥政策影响力的制度性窗口，后者便成为影响欧美人工智能监管模式差异的重要因素。

欧美公共部门所提供的游说渠道迥异。欧盟和美国的商业文化不同，美国的政企关系有机制化的游说渠道，欧盟更容易发展出“亲近但非正式”的关系。游说也并非欧盟政治的传统，在欧盟成立伊始，没有透明度登记和听证会制度，IBM、福特、通用汽车等美国公司的跨国经营和美国商业协会（American Chamber of Commerce）给欧盟带来了“美式游说”。[②] 欧洲企业对欧盟的游说有两种路径，一种是游说本国政府和理事会部长，该路径只提供符合本国利益的理由，比较容易成功；另一种是对欧盟机构的游说，该路径需要寻找符合“泛欧洲”的利益，是对整个欧洲的政策进行游说，只有能够提供欧盟所需的专业知识、欧洲包含利益信息和国内综合权益信息时才能够游说成功。[③] 美国的政治运行产生了悠久的利益集团传统，组织程度更高、资源更广的利益集团更能够影响政治结果。[④] 在选举政治中，利益集团可以通过提供政治捐款、协助开展竞选活动、动员选民投票等方式影响选举。[⑤] 赢得选举胜利的领导人也需要呼应利益集团关

---

① Alberto Bitonti，“The Role of Lobbying in Modern Democracy：A Theoretical Framework”，in Alb-erto Bitonti and Phil Harris，eds.，Lobbying in Europe：*Public Affairs and the Lobbying Industry in 28 EU Countries*（Palgrave Macmillan，2017），pp.17-30.

② David Coen，“The Impact of U.S. Lobbying Practice on the European Business Government Relati-onship”，*California Management Review*，Vol. 41，No.4，1999，pp.27-44.

③ Cornelia Woll，“Who Captures Whom？ Trade Policy Lobbying in the European Union”，in Coen David and Richardson Jeremy，eds.，*Lobbying in the European Union：Institutions，Actors and Issues*（O-xford University Press，2009），pp.268-288.

④ 李寿祺：《利益集团参政——美国利益集团与政府的关系》，载《美国研究》1989年第4期，第29-42页。

⑤ 同上。

注和选举承诺。在立法过程中，利益集团可以通过两院委员会和小组委员会及非正式的立法联盟影响立法。所以，美国不是没有监管苗头，而是出现监管苗头的时候会被科技公司压制和分散——美国政府与企业围绕监管问题的博弈呈现出“钟摆”式的动态特征。

欧美私营部门的游说能力也大相径庭，技术能力和联合能力是两大影响因素。鉴于人工智能技术和基础设施的垄断性，企业对技术的垄断能力对企业游说能力至关重要。在美国科技巨头繁多而欧盟企业多为初创公司的背景下，美国私营部门的游说能力远超欧洲企业，不仅在国内有雄厚的游说基础，对欧盟的跨国游说也形成了相对固定的代理集团和游说模式。[①] 在资金方面，美国科技公司在国内和在欧盟的游说开支都名列前茅。据Corporate Europe统计，大型科技公司在欧盟的游说力量从每年9700万欧元增至1.13亿欧元，谷歌、亚马逊、Meta等美国科技巨头位居前十。[②] 在人员方面，美国科技公司开始在国内和欧盟大量聘用专职人员，也有专职的游说机构为其服务。在美国国内，以OpenAI为例，该公司首席执行官出席国会作证时就聘请了律师指导作证，希望聘请一位美国国会领导，每年为该职位预算23万至28万美元。[③] 在欧盟，美国科技公司有信息技术产业委员会（ITI）、计算机和通信行业协会（CCIA，包括谷歌和亚马逊在内）以及商业软件联盟（BSA，包括微软和IBM在内）等游说集团。[④] 在话语权方面，美国科技公司对伦理和公共政策议程有很强的把控能力，通过资助研究所、智库为自己争取了话语权。生命未来研究所（Future of Life Institute）是美国科技公司争取话语权的典型案例。该组织自述为在美国和欧盟的非盈利组织，奉行长期主义的哲学思潮，Skype联合创始人贾恩·塔林（Jaan Tallinn）为联合创始人、特斯拉创始人兼首席执行官埃隆·马斯克（Elon Musk）也为该机构提供资助，该组织对美国人工智能政策实施和欧

① David Coen，“The Impact of U.S. Lobbying Practice on the European Business Government Relati-onship”，*California Management Review*，Vol. 41，No.4，1999，pp.27-44.

② “Byte By Byte：How Big Tech Undermined the AI Act”，Corporate Europe Observatory，November 17，2023，https：//corporateeurope.org/en/2023/11/byte-byte.

③ Hailey Fuchs and Brendan Bordelon，“AI Policy Yields a Goldmine for Lobbyists，”Politico，Nove-mber 4，2023，https：//www.politico.com/news/2023/11/04/ai-government-lobbying-00125378.

④ Max Bank，Felix Duffy，Verena Leyendecker and Margarida Silva，“The Lobby Network：Big Tech’s Web of Influence in the EU”，Corporate Europe Observatory and Lobby Control e.V. Brussels and Col-ogne，August 2021，pp.5-9.

盟《人工智能法案》都多有公开政策评论。[①]

企业的联合能力同商业利益和商业文化相关。美国企业在国内和在欧盟都有共同的商业利益并形成利益集团联盟，如向本国政府寻求产业补贴、优惠政策，向欧盟寻求降低关税、放松准入等。欧洲企业相对缺少一致的商业利益。从历史来看，这种商业利益不统一的情况在上世纪80年代和90年代欧洲一体化进入快车道的时期就有讨论，问题是关于欧洲各国的利益集团为什么没能联系成为欧洲法团主义，欧盟的政治一体化和经济一体化速度不一致就是其成因。[②]从法理来看，尽管存在欧洲经济一体化和单一市场，但一体化程度不足以完全打破国家政策对企业形成的无形壁垒。在人工智能这种成员国和欧盟权能不甚分明的领域，欧盟不能保证清除单一市场的碎片化，欧洲企业在母国和其他成员国的贸易同美国企业跨国经营的处境无二。

（三）产业生态影响

产业生态是技术发展的土壤，针对技术的治理需要考虑产业生态的影响。欧美存在差异性的产业生态，代表私营部门发展情况的技术生态、代表公共部门执行垄断与反垄断的政策生态、代表市场统一程度和规模的市场生态，这三者体现了公私部门的互动和偏好，共同作用于不同治理模式的形成。私营部门发展良好、反垄断政策缺失、市场统一，则监管阻力大，反之则更容易出台监管措施。

从技术生态看，欧盟行业力量相对薄弱，美国的公司规模、私营部门融资和产研转化能力都优于欧盟。从公司规模来看，欧盟不像美国一样拥有诸多科技巨头，而是在人工智能领域以中小企业为主。美国科技公司多为垄断发展，初创公司往往会被科技巨头收购，如微软收购OpenAI、谷歌收购Deepmind。2022年全球人工智能领域前五的并购和收购案有4个为美国公司，总金额达480.86亿美元。[③]而欧盟从事人工智能的科技公司少有大公司垄断的情况，更常见的是中小企业对人工智能技术进行应用。从

① “General Purpose AI and the AI Act”, Future of Life Institute, May 2022, https://artificialintelligenceact.eu/wp-content/uploads/2022/05/General-Purpose-AI-and-the-AI-Act.pdf.

② Wolfgang Streeck and Philippe C.Schmitter, “From National Corporatism to Transnational Pluralism: Organized Interests in the Single European Market”, *Politics & Society*, Vol.19, No.2, 1991, pp.133-164.

③ Nestor Maslej et.al, “Artificial Intelligence Index Report 2023,” p.185.

私营部门融资情况来看，美国初创公司的创新条件相较于欧盟要更为优越。Disfold2022年的数据显示，在全球30家融资最充足的人工智能初创公司中，有16家位于美国，9家在中国，3家在英国，1家在加拿大，1家在新加坡。[①] 同样根据麦肯锡2016年和2018年的数据，2016年，欧洲仅吸引了全球风险投资和企业融资的11%，另外50%的资金投入到了美国公司，其余资金流向了亚洲（主要是中国），而2018年只有4家欧洲公司进入全球人工智能初创公司100强。[②] 美国对人工智能的累计私营部门投资和年私营部门投资也远超欧洲。根据斯坦福大学《人工智能指数2023》，2013-2022年全球前15位人工智能投资国家中美国占据首位，共投入2489亿美元，欧盟成员国（德国、法国、西班牙）以及英国共投入363.7亿美元。[③] 从产研转化能力来看，美欧都有良好的研发力量，但美国的转化能力优于欧盟。美欧都有各自的研究机构和研究型大学，也有良好的公私合作体系，如美国国家基金会同研发机构的合作、欧洲“地平线计划”。在出版及影响力方面，美欧也不相上下。根据《人工智能指数》，美欧的出版论文、会议论文、存储论文数量都比较接近，全球占比均在10%-20%。[④] 不过根据专利引用量，美国情况更优：2014年至2018年间，北美地区人工智能专利引用量占全球人工智能专利引用量的60%以上。[⑤]

从政策生态看，欧盟有反垄断的政策传统，美国的反垄断政策一直执行不力。人工智能具有自然垄断的属性，大模型表现尤甚，也已经有诸多文献讨论人工智能对反垄断的影响。[⑥] 究其垄断成因，一共有三点。

① EU AI Governance，“Argument Against Impact：EU Is Not an AI Superpower”，Effective Altruism Forum，January 31，2022，https://forum.effectivealtruism.org/posts/suyb4vC75Wo9EKgyu/argument-against-impact-eu-is-not-an-ai-superpower#fnwhwj8qwuz3b.

② 同上。

③ Nestor Maslej et.al，“Artificial Intelligence Index Report 2023”，pp.190.

④ Nestor Maslej et.al，“Artificial Intelligence Index Report 2023”，pp.20-68.

⑤ EU AI Governance，“Argument Against Impact：EU Is Not an AI Superpower”，Effective Altruism Forum，January 31，2022，https://forum.effectivealtruism.org/posts/suyb4vC75Wo9EKgyu/argument-against-impact-eu-is-not-an-ai-superpower#fnwhwj8qwuz3b.

⑥ 讨论人工智能对反垄断的影响、人工智能企业的联合策略可参见：Shin-Shin Hua and Haydn Belfi-eld，“AI & Antitrust：Reconciling Tension Between Competition Law and Cooperative AI Development”，*Yale Journal of Law & Technology*，Vol.23，2021，pp.415-550；Václav Šmejkal，“Three Challenges of Arti-ficial Intelligence for Antitrust Policy and Law”，*Intereulaweast*，Vol.8，No.2，2021，pp.97-118；Suzanne R-ab，“Artificial Intelligence，Algorithms and Antitrust”，*Comparative Law Journal*，Vol.18，No.4，2019，pp.141-150。

其一，人工智能前期投入成本高，生产与投资沉没性明显，[①] 数据、芯片、算力、网络等基础设施要求高，资金、人才需求大。其二，人工智能的成本劣加性明显，虽然沉没成本高，但投入应用后的边际成本会降低，随着用户数量的增多可以不断有数据反馈，从而可以在更低成本的前提下改建人工智能算法和模型。[②] 其三，人工智能容易产生数据、资金、技术等壁垒，形成垄断经营，尤其在达到一定规模后，市场上的少量供应商在提供等量服务时反而比大量供应商成本更低。[③] 鉴于人工智能的自然垄断属性，反垄断政策就构成了人工智能发展的重要政策生态环境。

欧洲的反垄断制度存在执法导向和新自由主义（Ordo-liberalism）的特征。[④] 在具体政策表现中，欧盟一贯有执行反垄断政策的传统，《欧洲联盟运行条约》（Treaty on the Functioning of the European Union）第101条就反垄断进行规定。人工智能在欧盟属于广义的数字政策，欧盟的《通用数据保护条例》（General Data Protection Regulation）、《数字市场法》、《数字服务法》都对反垄断进行了规定。对于人工智能这样具有自然垄断属性的行业，欧盟内部也有很多声音支持“事前监管”。[⑤] 美国当前的反垄断制度存在司法导向和偏向大企业的特征。美国虽然有《谢尔曼法》（Sherman Act）等传统反垄断法案，以及《平台竞争和机会法案》（Platform Competition and Opportunity Act）、《终止平台垄断法案》（Ending Platform Monopolies Act）等针对新兴数字经济的反垄断法案，但一直缺乏执行力。[⑥] 自20世纪90年代美国诉微软之后，美国对科技巨头的垄断行为“睁一只眼闭一只眼”实际上已经成为了不成文的规定。[⑦] 近年来虽然在官方有2015年的美国联邦贸易委员会（FTC）诉亚马逊案、2020年的FTC诉脸书案，在民间有反对

① 许丽：《论通用人工智能治理中管制与反垄断的协同》，载《上海政法学院学报（法治论丛）》2024年第1期，第117-132页。

② 同上。

③ 同上。

④ Manuel Wörsdörfer, “Ordoliberalism 2.0: Towards a New Regulatory Policy for the Digital Age”, *Philosophy of Management*, Vol. 19, 2020, pp.191-215.

⑤ Christian Djeffal, Markus B. Siewert, and Stefan Wurster, “Role of the State and Responsibility in Governing Artificial Intelligence: A Comparative Analysis of AI Strategies”, *Journal of European Public Policy*, Vol.29, No.11, 2022, pp.1799-1821.

⑥ Anu Bradford, *Digital Empire: The Global Battle to Regulate Technology*, pp.1-40.

⑦ Anu Bradford, *Digital Empire: The Global Battle to Regulate Technology*, p.21.

科技巨头垄断的声音出现，[①] 但美国自身的国家安全需要就决定了无法对涉及到军民两用的先进技术真正采取反垄断措施。美国国防部是其中的重要推手，美国可以获得科技巨头的技术创新支持，国防部也有制造无人机武器相关的订单直接交给垄断企业（如Google的Maven项目、亚马逊的JEDI项目），对垄断企业进行拆分存在创新、采购、生产工具分配的一系列质疑。[②]

从市场生态看，欧盟是需要通过制定规则建立单一市场，美国是对本身存在统一市场施加额外限制。具体而言，欧盟的统一人工智能市场仍然处于形成阶段，打造单一市场可以实现欧盟公私部门的“双赢”；美国国内的人工智能市场统一且规模大，海外市场是美国私营部门的重要目标。对于欧盟而言，其主要的人工智能公司分布在各成员国之中，而成员国之间的贸易又占欧盟贸易总量半数以上，因此通过法律和政策手段推进人工智能统一市场的形成有利于扩大市场规模、减少贸易规约，更加符合欧洲人工智能发展的利益。因此，欧盟承担了推进形成人工智能产业统一监管框架的主要责任，因此其在推进法律和政策时受到来自企业的阻力较小。对于欧盟企业来说，统一规则有利于其在欧盟各成员国内自由流动，具有促进作用。其争议和阻力只是如何监管、按照谁的利益监管的问题，而不是是否需要监管的问题，需要监管以促进单一市场发展是欧盟及其成员国的共识。美国本身就存在统一的国内市场，联邦层面的规则和州层面的规则更多是具体规定的差别，而不像欧盟一样涉及国家主权、关税、文化等显著差异。尽管加利福尼亚州等地方对于人工智能的监管和使用已经走在前列，但地方和中央仍然是相互平衡的关系，行业自治的立法基调不会相差甚远。[③] 而且，美国科技公司不仅关注国内市场，其更重要的属性是跨国公司，拓展海外市场是企业发展的重要目标，不管是针对欧盟的《人工智能法案》还是针对美国国内的总统行政令，科技公司都不希望有严格的

① 江山：《美国数字市场反垄断监管的方法与观念反思》，载《国际经济评论》2021第6期，第26-55页。

② Dakota Foster and Zachary Arnold，“Antitrust and Artificial Intelligence：How Breaking up Big Te-ch Could Affect the Pentagon’s Access to AI”，Center for Security and Emerging Technology Issue Brief，2020.

③ Brendan Bordelon，“As States Move on AI，Tech Lobbyists Are Swarming in”，Politico，September8，2023，https：//www.politico.com/news/2023/09/08/tech-lobby-state-ai-efforts-00114778.

监管措施出台，还试图通过“主动呼吁监管”的方式为公司谋得监管先机。[①] 在此背景下，美国对人工智能监管的阻力就更大，制定规则对企业带来的价值增量小于其成本。

表2：欧美人工智能治理模式差异的影响因素

	政策传统		利益集团		产业生态		
	监管强度	权力维度	游说渠道	游说能力	行业力量	反垄断政策	单一市场
欧盟	强	规范	少	弱	弱	强	碎片
美国	弱	技术	多	强	强	弱	完整

资料来源：笔者自制

四、结语

随着人工智能日新月异的发展，“科林格里奇困境”再度凸显。如何统筹安全与发展，平衡监管与创新，不仅是各国国内决策者面临的一道难题，也是全球治理领域一个迫切需要应对的全新命题。由于在政策传统影响、利益集团影响以及产业生态影响等方面的差异，欧美在人工智能治理实践中采取了不同的路径与模式。总体上，欧盟采取的是基于权利和风险的“硬监管”模式，侧重法律层面的人工智能治理框架，发挥其“规范性力量”。美国采取的则是基于市场与技术的“软监管”模式，鼓励“行业自治”和“自我监管”。欧美在人工智能治理模式上的差异也成为美欧贸易与技术理事会（TTC）等跨大西洋平台协调的重点之一。面对欧盟《人工智能法案》带来的先发优势，美国面临“两难困境”，既需要维护其科技霸主地位，又希望维护其在规则和标准方面的影响力。

在人工智能治理上，中国面临着与美国类似的困境。一方面，在中美战略竞争与美国对华技术遏制的背景下，中国需要与美国竞争人工智能技术主导权，保持在技术创新上的领先地位；另一方面，随着欧盟《人工智能法案》的出台及其可能的溢出效应，中国又需要与欧盟竞争人工智能国际治理的规则制定权。因此，人工智能治理对中国而言不仅涉及技术之

① Michelle Toh and Yoonjung Seo，“OpenAI CEO Calls for Global Cooperation to Regulate AI”，C-NN，June 9，2023，https://edition.cnn.com/2023/06/09/tech/korea-altman-chatgpt-ai-regulation-intl-hnk/index.html.

争，也涉及规范之争。在人工智能治理的“技术—规范”之争问题上，欧美选择了不同的治理范式。实际上，欧美两种治理模式都是根据各自特殊情况探索出的合适道路，两者之间并不存在根本的冲突。

中国人工智能产业已经进入规模化应用的爆发期，又具备欧盟缺乏的统一大市场，因此具备调和欧美两种治理模式的潜力，找到一条平衡的人工智能治理道路。平衡监管与创新，兼顾技术与规范应该是中国未来人工智能治理的方向。一方面，随着人工智能技术竞争的加剧，国家选择何种治理模式将深刻影响人工智能技术的发展。在中美争夺技术主导权的背景下，人工智能为中国提供了驱动经济增长和实现技术升级的机会，因此，中国的人工智能治理首先应该激励技术创新与进步，避免过度和全面的监管束缚产业创新活力。另一方面，以欧盟《人工智能法案》为标志，围绕人工智能的规则与标准之争日趋白热化，国际社会也在推进人工智能的全球治理。中国作为人工智能强国，应积极参与人工智能的全球治理，为建立人工智能领域的多边国际机制或者国际组织发挥建设性作用。

人工智能全球治理机制复合体构建探析

桂畅旎

摘　要：人工智能具有内生安全、应用安全、国家安全以及地缘政治安全风险，并基于各类风险形成不同的治理路径，进而体现了机制复杂性的特征，导致人工智能全球治理困境加剧。人工智能全球治理困境的根源在于治理机制未能反映人工智能安全风险特点；治理格局未能反映人工智能行为体力量变化以及互动关系；治理主导理念未能反映人工智能技术发展趋势等。将机制复合体理论引入应对人工智能全球治理困境，构建人工智能全球治理机制复合体，基于分层，实施模块化的动态治理；基于中心，依赖多主体开展协同治理；基于信任，构建大国协调的责任治理。

关键词：人工智能；全球治理；治理困境；机制复杂性；机制复合体

作者简介：桂畅旎，中国信息安全测评中心高级工程师。

一、引言

人工智能作为当前全球最具突破性的技术之一，被世界各国决策者视为重要的战略资源，进而通过积极的政策扶持，推进人工智能赋能发展，以增强国家竞争力。然而，人工智能安全风险引发广泛担忧，从散布虚假信息到助推认知战，从加剧社会失业到干扰社会秩序，从赋能军事化运用到导致自主武器失控问题等，呈现出各类风险相互交织的复杂局面。区别于一般新兴技术，人工智能具有涌现性、应用广泛性、军民两用性、影响外溢性等特点，这些特点使得人工智能成为国际治理的重要议题。

目前国际上普遍认为人工智能全球治理始于2016年。[①] 自此以后，众

感谢匿名评审专家和编辑部对本文提出的意见和建议，文责自负。

① Lewin Schmitt，“Mapping Global AI Governance：A Nascent Regime in a Fragmented Landscape”，*AI and Ethics*，Vol. 2，2022，pp.303-314.

多国际组织开始认真对待人工智能相关政策议题，包括联合国、世界经济论坛和亚太经合组织等，均将人工智能治理列入政治议程，并从最早的部长级逐步提升到当前的国家领导人级别讨论的议题。受人工智能技术本身安全特性、各国人工智能治理战略考量以及传统技术治理惯性等影响，人工智能全球治理机制复杂性（Regime Complexity）的特点凸显，不同倡议之间相互作用、相互影响进而展现出机制复合体（Regime Complex）的特征。

当前人工智能全球治理步入十字路口。建构有效的人工智能全球治理体系不仅需要捋清人工智能的基本安全风险及特征，还需要回应人工智能治理面临的挑战。本文从风险分析出发，回应当前治理困境，并引入机制复合体理论构建人工智能全球治理体系，以期为该领域提供一个新的分析视角。

二、概念界定与研究现状

（一）概念界定

人工智能全球治理与人工智能全球治理机制复合体是本文主要概念。人工智能全球治理是全球治理在人工智能议题上的应用和延伸，涉及各方为应对人工智能安全风险而共同制定的一系列行为预期的规范、规则、标准、程序和执行机制。人工智能全球治理的基本目标在于确保人工智能生态系统的透明度、问责制、可解释性、包容性等，最大限度减少人工智能的风险和潜在不利影响。为达成这一目标，各国政府、技术社群、私营企业、国际组织等主体基于治理惯性、占有资源等要素，构建了一个集倡议、原则、法律、条约等多方面因素于一体的人工智能全球治理复杂系统，进而形成机制复合体。人工智能全球治理机制复合体是各种机制之间的松散耦合组成的机制复合体，[①] 根据不同的议题构建包括标准、规范、规则、条约、法律等不同层次的治理机制。

（二）研究现状

学术界高度关注人工智能全球治理进程，已有研究主要包括以下方面：

一是探讨人工智能全球治理机制。治理机制是为应对风险而采取的制

① 鲁传颖、约翰·马勒里：《体制复合体理论视角下的人工智能全球治理进程》，载《国际观察》2018年第4期，第67-83页。

度安排。在人工智能全球治理机制的选择上，耶鲁大学政治学教授艾伦·达福（Allan Dafoe）提出了共同监管、强制自我监管和元监管共同存在的混合治理模式。[①] 哈佛大学伯克曼·克莱因互联网与社会中心的执行主任乌尔斯·加塞（Urs Gasser）等提出了适应性治理，认为人工智能治理需要从“命令和控制”措施转向更灵活的方法，要求对法规和政策进行迭代调整和改进。[②] 瑞典乌默奥大学教授弗吉尼亚·迪格南 （Virginia Dignum）借鉴了“负责任研究与创新”（Responsible Research and Innovation，RRI）的方法，要求基于人类基本原则和价值观开发智能系统。[③] 基于人工智能技术的复杂性，有学者提出分层治理的模式。乌尔斯·加塞等从互联网治理结构的发展和演变经验中提出了人工智能治理的层次模型，即技术层、伦理层以及社会与法律层。技术层主要是对算法和数据的治理，具体手段包括数据治理、算法问责以及标准；伦理层主要基于道德的治理，手段包括标准和原则；社会和法律层主要是利用规范、监管和立法等手段进行治理。还有一些学者从互联网治理、太空、航空安全、军控等领域的经验中汲取教训，为人工智能治理提供启示。[④]

二是关注人工智能全球治理困境及应对举措。随着人工智能治理的发展，不少学者关注到人工智能治理面临的挑战。鲁传颖等认为大国之间的竞争导致三重困境，分别是：人工智能技术本身的安全复杂性导致国家间信任不足，阻碍了全球范围内形成成熟有效的人工智能治理模式；霸权国的针对性竞争战略阻碍了人工智能全球治理的操作空间；人工智能全球性治理机制的缺失削弱了安全困境缓解的可能性。鲁传颖等进而提出多方参与性、时间敏感性和反馈迭代性为核心特征的敏捷治理模式以破解人工智能治理难题。[⑤] 薛澜认为人工智能治理面临五项挑战，分别是：技术发展

① Dafoe Allan，“AI Governance：A Research Agenda”，Centre for the Governance of AI Future of Humanity Institute University of Oxford，July 2017，p.42.

② Urs Gasser and Virgilio A.F. Almeida，“A Layered Model for AI Governance”，IEEE Internet Computing Vol. 21 Issue. 6 ，2017，pp. 58-62.

③ Virginia Dignum，*Responsible Artificial Intelligence：How to Develop and Use AI in a Responsible Way*，（Springer Cham，2019），p. 6.

④ James Butcher and Irakli Beridze，“What is the State of Artificial Intelligence Governance Globally？”*The RUSI Journal*，Vol. 164，Nos（5-16），November 2019，pp.88-96.

⑤ 鲁传颖、张璐瑶：《人工智能的安全风险及治理模式探索》，载《国家安全研究》2022年第4期，第84-100页。

与治理体系步调不一致、企业与政府信息不对称、滥用与风险规制成本不对称、全球治理重叠和矛盾、地缘政治风险，因此提出加强安全和技术，引进敏捷治理，鼓励企业自我规制，加强国际治理，解决地缘政治问题等五方面举措。[①]

三是引入机制复合体理论分析人工智能全球治理态势。“机制复合体”作为当前全球治理的重要机制，已运用于应对气候变化、医疗健康、知识产权和网络空间[②]等全球议题。该理念最早由卡尔·罗斯特兰（Kal Raustiala）和戴维·维克多（David Victor）在2004年提出，意即治理某一特定问题领域的一系列部分重叠且非等级制的制度组合。[③]凯伦·奥尔特（Karen J. Alter）和索菲·梅尼尔（Sophie Meunier）则认为是“不按等级排列的，存在嵌套、部分重叠和平行的国际制度”。[④]罗伯特·基欧汉（Robert Keohane）将机制复合体描述为“一体化”和“碎片化”的中间状态，认为其成因主要有利益分配、不确定性和制度联系。[⑤]随着人工智能国际制度的发展，有学者将机制复合体运用于人工智能全球治理方面的研究。牛津大学皮特·西洪（Peter Cihon）认为人工智能全球治理可以被视为一种机制复合体，并从政治权力、效率支持和参与、中央集权制度的缓慢与脆弱、广度与深度的困境、选择法院、政策协调六个因素进行对照考虑。[⑥]斯德哥尔摩大学霍纳斯·塔尔贝格（Jonas Tallberg）认为目前人工智能全球治理的结构符合对机制复合体的描述，通过这一理论视角来看待人工智能治理，可以揭示人工智能治理的宏观属性，并可以开辟新的研究途径。[⑦]剑桥大学国王学院特约研究员马蒂亚斯·马斯（Matthijs Maas）将

① 赵珊：《薛澜：人工智能面临治理挑战》，载《中国经济时报》2024年3月25日。

② 相关论文见：约瑟夫·奈：《机制复合体与全球网络活动管理》，载《汕头大学学报（人文社会科学版）》2016年第4期，第87-96页；郎平、陈琪琪：《网络空间国际治理的机制复杂性及其影响》，载《同济大学学报》（社会科学版）2023年第6期，第47-59页。

③ Kal Raustiala and David G. Victor, “The Regime Complex for Plant Genetic Resources”, *International Organization*, Vol. 58, No. 2, 2004, pp. 277-279.

④ Karen J. Alter and Sophie Meunier, “The Politics of International Regime Complexity”, *Perspectiveson Politics*, Vol. 7, No. 1, 2009, p.13.

⑤ Robert O. Keohane and David G. Victor, “The Regime Complex for Climate Change”, *Perspectives on Politics*, Vol. 9, No. 1, March 2011, pp. 7-23.

⑥ Peter Cihon, et al., “Fragmentation and the Future: Investigating Architectures for International AI Governance”, *Global Policy*, Vol. 11, No. 5, November 2020, pp. 545-556.

⑦ Jonas Tallberg, et al., “The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research”, *International Studies Review*, Vol.25, No. 3, 2023, p.3.

人工智能全球治理视为机制复合体理论的“天然候选者”，认为利用该理论可超越技术本身，将制度在“碎片化”机制中以富有成效和非冲突的方式与其他机构或制度联系起来。[①]

上述文献对人工智能国际治理进行了丰富的研究，为理解当前治理的复杂现实奠定了坚实的基础。既有研究也肯定了机制复合体理论与人工智能全球治理的契合度，但对于如何利用机制复合体建构人工智能全球治理体系的路径却少有着墨，这也是本文尝试回答的核心问题。

三、人工智能安全风险

风险是塑造安全的起点，人工智能治理离不开对风险的识别。人工智能的包容度与统摄力使其具备主导技术发展和推动社会形态转变的基本潜质，[②]同时也因其通用性、赋能性和自适用性等特点，使其与其他技术风险存在显著区别（见表1）。一是技术不可控性强。人工智能的自适应性以及自我迭代性强，能够在短时间内积聚变化，再加上人工智能的“黑箱”性质，其不稳定性完全超出人类的控制。二是治理资源非国家垄断。科技企业是人工智能技术资源的主要保有者，同时也是安全解决方案的主要提供者，国家不再垄断人工智能治理资源，不得不依赖企业的专业以及信息。三是安全威胁影响泛化。人工智能引发的风险已不限于特定领域、特定范围，而是威胁全人类生存。不再是单个国家思考如何应对另外一个国家的安全威胁问题，而是国家群体思考如何合力应对共同的安全威胁问题。[③]本文主要从内生安全、应用安全、国家安全和地缘政治安全四个层面来分析人工智能安全风险。

（一）人工智能内生安全（AI Safety）

人工智能治理具有典型的技术治理特征，即关注系统的自身安全。大多数技术治理议题是通过制定标准、规范的方式来应对技术缺陷问题，通

① Matthijs Maas, *Artificial Intelligence Governance under Change Foundations, Facets, Frameworks*, Ph.D. dissertation, UCPH, 2021.

② 阙天舒、张纪腾：《人工智能时代背景下的国家安全治理：应用范式、风险识别与路径选择》，载《国际安全研究》2020年第1期，第4–38页。

③ 秦亚青：《全球治理失灵与秩序理念的重建》，载《世界经济与政治》2013年第4期，第4–18页。

表1：人工智能与其他技术风险比较

类别	特点	
	人工智能	一般技术
技术可控	不可控	相对可控
治理资源	企业	国家
安全影响	泛化	限定

资料来源：笔者自制

常情况下并不需要政治介入，也不会成为国际治理议题。但由于人工智能自身安全特性以及潜在的外溢性风险，使得其内生安全问题不仅受到各国政府高度关注，而且成为人工智能治理领域国际对话与交流的核心议题。

人工智能内生安全源于其框架、组件、数据、算法、模型等存在的脆弱性。框架、组件缺乏透明性；数据的丢失和变形、噪声数据的输入；算法的偏见和歧视；模型易被窃取和被植入后门等都属于人工智能的内生安全问题。[①] 人工智能内生安全风险主要源于：一，人工智能技术自身的脆弱性。这可追溯到人工智能在设计之初普遍未考虑相关的安全威胁，导致人工智能算法的判断结果可能会产生与人类目标不完全一致的结果。例如，自动驾驶汽车带来的撞车事故，主要原因是系统误解了之前在测试过程中未经历过的独特环境条件。二，人工智能的自主性和独立逻辑将会产生一定程度的不可预测性。在最新一轮由大数据分析推动的人工智能发展浪潮中，人工智能依赖于机器学习算法可以从大量数据中学习，并在没有人类指导的情况下做出决策，这与早期依赖预编程规则来执行重复任务的算法不同，但也容易造成系统尚未训练处理的意外情况，即系统在相同输入的前提下也可能会表现出截然不同的行为。人机交互中的不确定性会导致人工智能系统出现对用户构成安全隐患的意外行为。[②]

人工智能固有的不透明性和不可预测性给治理带来了技术挑战。一方面，复杂算法的不透明性限制了人工智能系统的可解释性，使得治理主体难以在具体政策中制定具体的治理目标。另一方面，人工智能决策

① 姚期智，《人工智能目前的算法有较大的脆弱性和不稳定性》，新浪财经，2020年10月22日，https://finance.sina.cn/hy/2020-10-22/detail-iiznctkc7063082.d.html。

② Araz Taeihagh，“Governance of Artificial Intelligence”，*Policy and Society*，2021，Vol. 40，No. 2，pp. 137-157.

的不可预测性使得人类难以对其进行控制，这也就意味着治理主体难以对软件缺陷进行问责。

（二）人工智能应用安全（AI Security）

人工智能技术自身的不可控性以及不透明性难以解决，在应用过程中呈现出泛化和滥用的趋势，随之产生人工智能的应用安全问题。人工智能作为基础平台性技术，犹如“电力”一样，赋能于社会各个方面，但其内生安全问题也随之被放大。具有恶意使用目的的客体针对人工智能基础环节，开展违背设计者初衷的行为与活动，引发数据安全问题以及网络攻击和算法操纵等安全威胁。基于人工智能的网络攻击最为突出的案例就是对抗样本攻击，其实质是利用算法的缺陷进行攻击。

不同于常规军民两用技术的高度复杂性和资本密集性，人工智能作为一项通用型技术，其低成本以及易传播的特点使其成为全新的力量投射手段，[①] 这也导致其安全风险泛在化更为凸显。从应用主体方面来看，人工智能技术对于应用者的门槛在不断降低，主体更为多元，应用领域更广泛，潜在的安全风险也就更为突出。从应用领域来看，人工智能的广泛应用衍生出技术滥用、数据安全、网络安全、隐私保护等安全挑战。例如，人工智能在采集、使用和分析海量数据过程中，容易发生隐私泄露、数据篡改等风险；人工智能运用到无人驾驶、医疗诊断等领域，可能引发权责边界模糊等问题；此外人工智能系统的自治性质也可能引发人类自主性和决策控制的丧失问题。

人工智能应用风险并非单一、直线的，而是交织交融的，这也使得人工智能治理不仅是技术问题，还涉及到政治和社会经济问题。这不仅需要从技术层面解决滥用问题，还需从社会治理和伦理规范层面进行引导。

（三）人工智能国家安全风险

人工智能作为一项强赋能型技术，同样应用于国家安全领域。特别是在新一轮发展浪潮下，人工智能与国家安全的嵌入程度更为深入，因而产生一系列国家安全风险。2017年7月，哈佛大学肯尼迪学院贝尔福科学与国际事务中心发布《人工智能与国家安全》报告，认为未来人工智能有可能成为与核武器、飞机、计算机、生物技术不相上下的变革性国家安全

① 《AI治理的基本要点》，国际货币基金组织，2023年12月，https://www.imf.org/zh/Publications/fandd/issues/2023/12/POV-building-blocks-for-AI-governance-Bremmer-Suleyman。

技术。[①] 人工智能国家安全风险主要体现在传统与非传统国家安全领域的智能化过程中。在军事层面，人工智能可用于构建新型军事打击力量，整合到指挥和控制，监视、情报和侦察，后勤和培训等模块，产生智能自主武器，这不仅可能颠覆传统战场态势，而且当与核武器结合，甚至可能对人类生存构成威胁。在社会层面，人工智能通过机器学习算法、聊天机器人和无人驾驶汽车实现数据分析、服务、制造和驾驶等领域的常规任务自动化，带来大面积的失业。人工智能革命还使复杂的恶意软件的出现与开发成为可能，增加了对民用和国防基础设施的网络攻击的威胁。

但正如亨利·基辛格博士（Henry Kissinger）所说，“与核武器领域不同的是，对人工智能的使用并不存在被广泛认同的禁令，也没有明确的威慑（或升级程度）概念”。[②] 人工智能获取门槛低，难以控制其应用范围，覆盖低政治领域到高政治领域，且还处于更新迭代过程中，其发展甚至对人类生存提出了挑战。2023年5月，数百名科学家和行业领袖发表了一份声明，宣称“减轻人工智能造成的灭绝风险应该与流行病和核战争等其他社会风险一起成为全球优先事项。”[③] 人工智能生态系统的跨境性质使得纯粹的国家监管制度效率低下且成本高昂。[④]

（四）人工智能地缘政治风险

人工智能的内生安全、应用安全以及国家安全风险具有较强的外溢性，导致地缘政治、战略考虑主导了各国政府对人工智能的监管或立法措施，进而引发地缘政治风险。当前世界主要大国普遍将人工智能资源视作战略性资源以及赋能国家实力的重要工具进行优先治理，[⑤] 在资源争夺上呈现出零和性、对抗性的特征。例如，美国为维持人工智能的技术优势，加大对人工智能基础资源的限制。一是强化算力控制，包括升级人工智能

① Greg Allen and Taniel Chan, “Artificial Intelligence and National Security”, Belfer Center for Science and International Affairs, Harvard Kennedy School, July, 2017, https://www.belfercenter.org/publication/artificial-intelligence-and-national-security.

② 亨利·基辛格、埃里克·施密特、丹尼尔·胡滕洛赫尔著，胡利平、风君译：《人工智能时代与人类未来》，中信出版社，2023年版，第186页。

③ Kevin Roose, “A.I. Poses ‘Risk of Extinction,’ Industry Leaders Warn”, The New York Times, May 30, 2023, https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html.

④ Lewin Schmitt, “Mapping Global AI Governance: A Nascent Regime in a Fragmented Landscape”, *AI and Ethics*, Vol. 2, 2022, pp.303-314.

⑤ 鲁传颖：《人工智能：一项战略性技术的应用及治理》，载《人民论坛》2024年第1期。

芯片出口禁令，全面禁止向中国销售英伟达、AMD及其他公司的先进人工智能芯片和半导体设备；二是针对核心算法进行保护，将核心算法和模型纳入《美国知识产权法》的保护范围，限制外国人使用美国的基础设施即服务（IaaS）开发大模型，实施算法阻断和模型垄断；三是限制重要数据流向，拜登签署了旨在“阻止外国实体获取大量美国人个人数据”的行政令，限制对手国家获取美国人和美国政府的敏感数据。地缘政治因素正逐渐侵蚀本应属于技术领域的人工智能发展与合作。

人工智能与传统国家安全领域的结合所带来的外部性越来越明显，特别是人工智能军备竞赛正在兴起。包括美国及其盟友在内的国家均在积极开发自主武器，广泛联合微软、谷歌等私营企业研发国防领域的人工智能技术，无人武器的智能化程度正在不断提升。被誉为人工智能之父的杰弗里·辛顿（Geoffrey Hinton）认为，“人类未来需要担心的不是机器的智能化提升，而是杀手机器人”。[①] 这也导致一些国家为了维持领先优势可能会容忍安全缺陷和治理投机，加剧既有的甚至产生新的冲突形势。[②] 此外，人工智能在赋能各国社会发展上极不平衡。发达国家往往拥有较强的技术实力和企业资源，为人工智能的发展做好了充足的准备；但广大发展中国家则缺乏相应的发展资源，其劳动力资源在这场以技术为核心的竞赛中不再占优势，造成发达国家与发展中国家的差距进一步拉大，新的“南北问题”显现。[③]

四、人工智能全球治理困境

人工智能全球治理需要回应人工智能安全风险带来的挑战。尽管当前针对人工智能安全风险的全球治理议程不断出现，但既有治理格局无法有效应对人工智能安全威胁的激增。佐治亚理工学院丹尼尔·希夫（Daniel Schiff）等在分析超过80篇来自政府和非政府组织制定的人工智能文件的基础上，提

① Joe Shute，“The‘Godfather of AI’on Making Machines Clever and Whether Robots Really Will Learn to Kill Us All?”，The Telegraph，August 26，2017，https://www.telegraph.co.uk/technology/2017/08/26/godfather-ai-making-machines-clever-whether-robots-really-will/.

② Olivia J. Erdélyi and Judy Goldsmith，“Regulating Artificial Intelligence：Proposal for a Global Solution”，In Proceedings of the 2018 AAAI/ACM Conference on AI，Ethics，and Society（AIES’18），Association for Computing Machinery，New York，December 2018，pp.95-101.

③ 李艳：《高度关注AI时代的“智能鸿沟”问题》，载《环球时报》2023年4月21日。

出人工智能伦理、政策和治理相关文件的成功与否主要取决于五个因素：与特定法律和政策的结合度，内容描述的具体化程度，对外共享及曝光度，执行过程的强制性程度，以及有效监督和快速迭代及持续更新，[①] 侧面揭示了当前人工智能全球治理手段和路径与治理目标、预期之间存在较大差距。

（一）当前治理规则未能反映人工智能安全风险特点

据不完全统计，全球活跃的人工智能安全治理倡议已高达50多项（见表2），治理主体涉及多边、多方甚至个体，治理议题囊括意识形态塑造、军事化管控、社会治理等多个方面，治理成果既有松散的磋商机制，也有超国家性质的立法，是目前全球参与主体最为广泛、涉及议题最为多样的治理领域之一。然而，现有的人工智能治理规则未能充分反映人工智能安全风险的特点，造成治理效果不佳、规则冲突等困境。

表2：当前全球主要人工智能倡议

组织形式	倡议
多边	欧盟人工智能委员会（CAI）《人工智能公约》；欧盟《人工智能法案》（AI Act）；七国集团广岛进程；联合国高级别人工智能咨询机构；自由在线联盟“人工智能与人权特别工作组”；二十国集团《负责任的以人为中心的人工智能治理框架》；欧洲刑警组织“人工智能问责原则项目”；经合组织（OECD）“人工智能政策观察站”；欧盟委员会“欧洲人工智能联盟”；北约“北约数据和人工智能审查委员会”
多方	世界经济论坛“人工智能治理联盟”；美国白宫/领先人工智能公司“人工智能联络小组”；卡内基国际事务伦理委员会“全球人工智能观察站”；Anthropic、谷歌、微软和OpenAI发起的前沿模型论坛；阿兰图灵研究所“人工智能标准中心”；互联网治理论坛“人工智能政策网络”；全球人工智能伙伴关系“生成式人工智能时代建立信任的全球挑战”；全球互联网反恐论坛“Hash共享数据库”；美洲开发银行“人工智能试点项目”；人工智能伙伴关系“公共政策项目”；巴黎和平论坛“利用人工智能改善公共政策”；信息与民主论坛“人工智能工作组”
国际组织	联合国高级别人工智能咨询机构；联合国教育、科学及文化组织（UNESCO）《关于人工智能伦理的建议》；联合国人权理事会NET决议；世界卫生组织（WHO）“卫生人工智能伦理和治理专家组”；联合国《全球数字契约》；《特定常规武器公约》缔约国“致命自主武器系统政府专家组”；世界贸易组织“数字技术与贸易”；联合国儿童基金会《生成人工智能：儿童的风险与机遇》；国际电信联盟“AI向善全球峰会”
政府引领	欧美贸易技术理事会（TTC）《可信赖人工智能和风险管理联合路线图》；英国全球人工智能安全峰会；金砖国家未来网络研究院“人工智能研究小组”；荷兰政府“军事领域负责任人工智能峰会”；中欧高级别数字对话

资料来源：笔者根据网上信息自制

① Daniel Schiff, et al.,“What's Next for AI Ethics, Policy, and Governance? A Global Overview”, In Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES '20), , February 2020, pp.153–158.

其一，治理规则笼统模糊与治理精细化要求的差距越来越凸显。人工智能全球治理规则过于笼统，专门性创新型倡议不足。一方面，人工智能技术发展已经超出治理部门的专业知识范围，导致政府传统监管资源不足，缺乏专业的技术与知识储备以管理人工智能带来的风险。另一方面，各国政府不愿过多让渡国家主权，对于建立专门的人工智能治理组织持谨慎态度。尽管存在人工智能全球伙伴关系（GPAI）、人工智能国防合作伙伴关系等专门规制，但是各国仍更倾向于在既有的、经过验证的治理框架基础上嵌入人工智能倡议。例如，七国集团、二十国集团、欧盟委员会、经合组织都是在既有机制上讨论人工智能议题。上述因素导致目前政府组织规制人工智能多是基于原则性的约束，主要围绕提升人工智能“可解释性”“公平性”“可问责”“透明度”等基本面展开，[①] 缺乏针对人工智能技术监管的特异性。[②]

其二，治理规则碎片化与治理专业性要求的分野越来越明显。从技术监管的生命周期来看，目前对于人工智能的安全治理仍然处于初期。基于不同治理经验和治理资源，国际组织、政府部门、私营部门、学术界和公民社会关注的议题以及路径偏好各异，导致议题之间出现了嵌套、重叠和平行的碎片化局面。[③] 虽然有学者认为碎片化的局面中可能存在自组织的趋势，并且不一定导致所谓的“选择法院”（Forum Shopping，是英美法术语，指当事人利用国际民事管辖权的积极冲突，从众多有管辖权的法院中选择一个最能满足自己诉讼请求的法院去起诉的行为）的困境，[④] 但是制度间的原则、规范和规则冲突，以及重叠的规定容易造成“条约拥堵”的情况，不仅增加合规成本，也极大削减了人工智能全球治理整体机制的权威性和有效性。

其三，国际机制的非约束性难以满足治理效用要求。当前人工智能全

① “Explainable Artificial Intelligence（XAI）”，DEFENSE ADVANCED RESEARCH PROJECTS AGENCY，https://www.darpa.mil/program/explainable-artificial-intelligence.

② Michael Guihot，et al.，“Nudging Robots：Innovative Solutions to Regulate Artificial Intelligence”，*Vanderbilt Journal of Entertainment and Technology Law*，Vol. 20，Issue. 2，385（2020），pp.385-456.

③ Butcher，J.，and Beridze，IJames Butcher and Irakli Beridze，“What is the State of Artificial Intelligence Governance Globally?”*The RUSI Journal*，Vol.164，Nos(5-/6)，November 2019，pp.88-96.

④ Peter Cihon，et al.，“Fragmentation and the Future：Investigating Architectures for International AI Governance”，*Global Policy*，Vol. 11，No. 5，November 2020，PP.545-556.

球治理制度多是不具有强制性的行为规范或政策建议，这难以约束政府和非政府行为体的活动。例如，联合国教科文组织（UNESCO）于2021年11月发布的《人工智能伦理建议》参与国家多达193个，但由于协议不具有约束力，只有不到四分之一的签署国与该机构合作实施了细化原则的政策工具。再如，由亚马逊、苹果、谷歌、Facebook、IBM和微软在2016年建立的人工智能合作伙伴关系组织（PAI）呈现出“高开低走”趋势，已有不少的成员和非政府组织退出，其主要原因是该组织未能发挥应有的效力。当国际规则无法充分发挥治理功能，国家只能绕过全球性规则而进行小范围的合作，所表现出来的恰恰是全球治理规则的低效度和不充分性。[①]

其四，人工智能技术迭代与治理滞后之间的矛盾难以弥合。人工智能技术发展迅速，迭代升级快，其扩散速度已经超出了此前的任何一项划时代技术，[②]人工智能技术的大规模实际应用只用了1年，相比之下，互联网普及用了7年，电力普及则用了20年，体现了人工智能技术超强的扩散速度。现有全球治理规则主要依赖于法规、标准等静态的治理规则，依靠冗长的谈判和协商，很难跟上人工智能技术快速迭代发展而带来的安全风险，这也导致技术发展与政策响应的时间差越来越大。“新技术所固有的扩散天性迄今为止使通过谈判做出限制的一切努力付诸东流，甚至连概念也未能形成”。[③]

（二）当前治理格局未能反映人工智能行为体力量格局变化

一方面，区别于其他前沿颠覆性技术，人工智能技术发展的核心驱动力量主要来自于全球性科技企业。少数企业控制着人工智能的系统资源，包括物理资源（如大型图形处理单元集群）、认知资源（接触尖端研究小组的机会）和信息资源（访问数据集的机会）。[④]在此基础上，企业成为应对人工智能安全风险的主力，通过联合非政府组织、学术团体等“自下而上”推动形成认知共同体。例如，人工智能合作伙伴关系组织（PAI）

① 秦亚青：《全球治理失灵与秩序理念的重建》，载《世界经济与政治》2013年第4期，第4-18页。

② 《AI式“技术扩散”：过去100年的10个“前车之鉴”》，“华尔街见闻”，2023年6月4日，https://wallstreetcn.com/articles/3690302。

③ 亨利·基辛格、埃里克·施密特、丹尼尔·胡滕洛赫尔著，胡利平、风君译：《人工智能时代与人类未来》，中信出版社，2023年版，第200页。

④ Michael Veale, et al., “AI and Global Governance: Modalities, Rationales, Tensions”, *Annual Review of Law and Social Science*, Vol.19, 2023., pp.255-275.

通过诸如《安全关键型人工智能》白皮书等出版物，为创建可信赖的人工智能系统提供了全面的框架。[①] 企业的最佳实践能够快速跟进技术发展的步伐，相比于政策法规能够及时进行修订和调整，对于促进人工智能治理中的道德、非歧视做法具有重要作用。但是这类自下而上的塑造力量并未在人工智能全球治理体系中找到合适位置，掌握核心治理资源的政府与掌握技术与信息的企业和行业团体之间协调程度有限。无论是在政府内部还是在各种区域和国际论坛上，政企合作仍处于初级阶段，企业的治理实践未充分融入到政府间倡议中去。如何平衡好政府和企业的力量是构建人工智能治理体系需要回答的关键问题。

另一方面，虽然发达国家与发展中国家在人工智能资源占有和基础能力上差距显著，但是人工智能安全风险却在无差别影响着广大发展中国家，但“基于实力安排座位”的传统国家治理机制严重影响人工智能全球治理的公正性。目前人工智能的开发和应用仍然集中在欧洲、北美以及日本和韩国等发达经济体，非洲、拉丁美洲和亚洲大部分地区发展中国家在全球治理机构中的代表性不足，相关政策和活动也比较少。以人工智能伦理规制为例，目前美国以21部排名第一，欧盟（19部）紧随其后，英国（13部）和日本（4部）位居三、四。[②] 发展中国家在人工智能全球治理中缺位主要是由于缺乏治理资源和话语权，这反过来也会导致人工智能全球治理的议题主要反映发达国家的关切，发展中国家关注的发展议题、赋能议题由于代表性不足而难以在重要倡议中得到体现。

（三）当前治理理念未能反映人工智能技术发展规律

某种程度上来说，人工智能是一种超出现有监管或法律治理机制范围的新技术。[③] 但是，当前主导的治理理念仍然是基于西方惯有路径，这与人工智能技术发展的规律和趋势存在不适配。

传统安全至上的治理理念与人工智能技术发展规律不适配。人工智能

① Maral Niazi，“Conceptualizing Global Governance of AI”，Centre for International Governance Innovation，February 27，2024，p.6.

② Inga Ulnicane，et al.，“Governance of Artificial Intelligence：Emerging International Trends and Policy Frames”，in Maurizio Tinnirello ，“The Global Politics of Artificial Intelligence”（CRC Press.，2022），p.38.

③ Lewin Schmitt，“Mapping Global AI Governance：A Nascent Regime in a Fragmented Landscape”，*AI and Ethics*，Vol. 2，2022，pp.303-314.

往往被政治主体特别是霸权国视为“独占性”资源，以巩固自身的非对称优势。这种逻辑下，霸权国更倾向于遵循抑制技术扩散的传统政治逻辑，这违背了人工智能内生的外向扩散和升级演化的逻辑。例如，美国在人工智能资源管制上的对华打压引发了科技企业的不满和担忧，认为这样会阻碍技术的自然发展和创新。这也反映出传统的技术管控思维难以适应当前的技术发展，进而迟滞技术发展和社会进步就在所难免。

“少边主义”治理路径与人工智能技术开放性的内生驱动力之间的矛盾越来越凸显。美国正试图在全球范围内构建人工智能治理小多边机制，将自身倡导的价值理念融入到北约、经合组织、七国集团、人工智能全球伙伴关系（GPAI）以及美日韩、澳英美等多边机制中，试图构筑以美国为中心的人工智能治理的“少边主义”。然而人工智能技术的供应链和其所依赖的基础设施的价值链具有跨国跨界因素，人工智能应用产生的跨境外部性需要国际合作，人工智能开发需要通过跨国进程完成因而也需要跨境监管，[①] 排他性和竞争性的议题联盟可能成为合作的障碍。

西方价值观主导下的治理原则与人工智能发展的包容性特征不适配。目前世界主要大国竞相影响人工智能价值观对齐问题，意图以自身的意识形态训练人工智能。美欧等国均积极将自身价值理念嵌入人工智能治理规则里，例如在“以人为本”基本原则上，主要是以西方狭义的人权观念进行阐述，严重限制了人工智能技术的包容性特征。如何在底层价值观超越狭隘的西方语境，纳入更多的发展中国家的关切和诉求，建立更广泛、更公平、更聚焦于发展以及更具包容性的基本原则，将是考验人工智能是否能成为一项赋能全人类的变革性技术的基本考量。

五、以“机制复合体”理念构建人工智能全球治理体系

目前围绕人工智能的全球地缘政治竞争愈演愈烈，短时间内难以形成一个单一、一致和制度性的人工智能治理框架，全球人工智能治理处于且可能长期处于复杂的混乱状态。将机制复合体理论引入人工智能全球治理的分析视野，不仅从理论上有利于将行为体之间的互动和既有治理机制结

① Jonas Tallberg, et al., “The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research”, *International Studies Review*, Vol.25, No. 3, 2023, p.6.

合起来，从动、静两个维度对人工智能全球治理做出整体性分析，[①] 同时对于缝合碎片化的治理层级，应对人工智能全球治理困境具有重要作用。

（一）机制复合体基本特征及与人工智能全球治理的适配性

机制复合体主要有五个特征：一是主体多元性，包括国际组织、国家、社会组织、私营企业等主体，涉及国家间、公私间以及私人制度等形式。二是治理机制灵活，松散的连接机制能够容纳多种制度尝试，机制之间可能产生系统效应的相互作用，[②] 具有较强的灵活性。三是治理具有层次性。机制复合体是一个由若干机构和倡议组成的多边机制综合体，每个机构和倡议都涉及不同的成员集团，因此可基于对象进行分层治理。四是手段多样性。机制复合体理念能够将原则、政策、规范和法律结合起来，形成工具集，包括自上而下的国际法律规范以及自下而上的实践驱动影响。五是治理进程的持续性。机制复合体针对治理对象全生命周期进行治理，包括在部署前进行充分测试，部署后持续监控，不断评估技术发展的能力及其后果。[③]

机制复合体的基本特征使其能够更好地“实质契合”人工智能的风险特点和突破全球治理困境。

一是从理论框架来看，机制复合体的多中心、多模块的框架与人工智能全球治理特点适配性高。其一，机制复合体的多中心特性符合当前人工智能全球治理以联合国、经合组织、七国集团为主的多中心态势；其二，机制复合体的层次化特征能够解释当前人工智能全球治理中横向互动、纵向重叠的结构；其三，机制复合体存在的冲突互动或跨制度溢出等现象，与人工智能能力及用例的相互交织具有适配性。[④]

二是从功能角度来看，机制复合体的有机协作可有效改善人工智能全

① 郎平、陈琪琪：《网络空间国际治理的机制复杂性及其影响》，载《同济大学学报》（社会科学版）2023年第6期，第47-59页。

② Thomas Gehring and Benjamin Faude，“A Theory of Emerging Order within Institutional Complexes：How Competition among Regulatory International Institutions Leads to Institutional Adaptation and Division of Labor”，The Review of International Organizations，December 2014，p.4.

③ Emma Klein and Stewart Patric，“Envisioning a Global Regime Complex to Govern Artificial Intelligence”，Carnegie Endowment for International Peace，March，2024，https://carnegieendowment.org/2024/03/21/envisioning-global-regime-complex-to-govern-artificial-intelligence-pub-92022.

④ Matthijs Maas，*Artificial Intelligence Governance under Change Foundations，Facets，Frameworks*，Ph.D. dissertation，UCPH，2021.

球治理困境。其一，可有效改善碎片化问题。机制复合体中多元行为体根据自身的优势进行分工与协作，不断增强与其他制度的关系互动和协调整合程度，能够减少制度间“低效的重叠和不一致性”。其二，可提高治理效力。机制复合体的制度多样性有助于政策制定者将人工智能风险分解为可治理的具体模块，促使不同机构在不同的领域追求合作目标，以提升效力。其三，机制复合体能够充分结合软法和硬法实现综合治理。一方面，通过硬法确立精确的法律义务，提高违约成本，增强遵从性；另一方面，利用灵活的软法进行补充，鼓励集体学习，加强深入合作。[①]

三是从现有治理态势来看，人工智能机制发展趋势符合机制复合体的基本特征。其一，议题之间的互嵌融合趋势。目前，人工智能国际倡议之间的“联动”正在增多。[②] 如，七国集团广岛进程的核心条款均是沿袭经合组织的基本原则；欧美贸易技术理事会的“可信赖人工智能和风险管理联合路线图”则以欧盟《人工智能法案》为基础；美英《大西洋宣言》提出建立隐私保护技术的合作机制（PETs）等举措与七国集团如出一辙，表明各类倡议在沟通交流中出现趋同。其二，治理主体呈现出组合体趋势。在人工智能治理领域，政府、私营部门和学术组织的互动和融合不断增强，呈现出国家行为体和非国家行为体的组合体趋势。人工智能全球伙伴关系（GPAI）、经合组织的人工智能政策观察站（OECD.AI），世界政府峰会（WGS）主办的人工智能全球治理圆桌会议 （GGAR），以及联合国科教文卫组织关于人工智能的工作流程均是公私合作的典型。

（二）人工智能全球治理机制复合体的主要路径

人工智能全球治理机制复合体是当前人工智能全球治理机制有机互动、相互融合的结果，其主要路径体现在动态治理、协同治理以及责任治理三方面（如图1）。

1. 基于分层，实施模块化的动态治理

人工智能全球治理机制复合体主要针对人工智能安全风险特点实行分层、模块化的动态治理。

① Olivia J. Erdélyi and Judy Goldsmith，“Regulating Artificial Intelligence：Proposal for a Global Solution”，In Proceedings of the 2018 AAAI/ACM Conference on AI，Ethics，and Society（AIES’18），Association for Computing Machinery，New York，December 2018，pp.95-101.

② Huw Roberts，et al.，“Global AI Governance：Barriers and Pathways Forward”，*International Affairs*，September 2023，p.3.

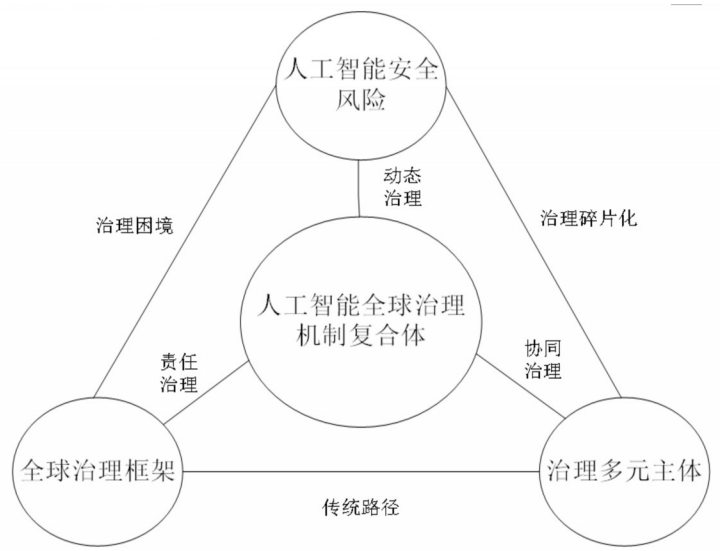

图1：人工智能全球治理机制复合体图

资料来源：笔者自制

一是基于人工智能安全风险的类别进行分层治理。与其他技术风险相比，人工智能安全风险决定了复合体需囊括微观的技术治理与宏观的政策治理，并进行分层治理：技术层治理主要是规制人工智能生态系统的基础即算法和数据，治理主体主要是掌握核心和底层技术的企业，往往依赖于制造商和软件设计师严格的产品责任制，通过实施“设计安全”原则，压实对制造缺陷和设计缺陷的责任。[①] 伦理层治理主要关注在技术层之上适用于所有类型的人工智能应用程序和系统的道德问题。一方面基于企业的自律要求，例如，谷歌、IBM、微软等企业提出的自愿性标准、指导方针和行为准则；另一方面需要中立性的国际组织基于事实准则进行非意识形态的治理，例如，电气和电子工程师协会（IEEE）建立的《人工智能设计的伦理准则》，提出了人权、福祉、问责、透明、慎用五大总体原则。社会和法律层面主要是国家以及国际组织等主体树立基本原则，分配监管人工智能责任的过程。在此过程中，国家以及国际组织主要通过硬法确立基本原则和精确的法律义务，同时通过各专家团体“涓滴”到技术层，影

① Carl Gahnberg，“What Rules？ Framing The Governance of Artificial Agency”，*Policy and Society*，Vol. 40，No. 1，June 2021，pp.1-21.

响具体操作。[①]

二是基于人工智能安全风险的迭代升级进行动态治理。有效的人工智能机制复合体需要针对人工智能安全风险全生命周期进行监督。在理念选择上，可引入敏捷治理模式以形成快速适应情景变化的动态治理范式，有效弥合人工智能快速变迁与政府监管相对滞后之间的矛盾。[②] 在实际操作中，构建事前预防与事后应对相结合的适应性治理机制，即在部署前需进行充分测试，引入测评、核查等政策手段，例如，美国国土安全部发布的《2024年人工智能路线图》就明确规定为人工智能系统建立严格的开发、测试和评估实践的标准；[③] 及时跟踪技术和应用的发展事态，尽早分析并识别各种安全风险的严重程度，进而动态调整和快速响应。[④] 此外，还需要针对治理政策的有效性进行定期评估和实时调整，以最大限度适应人工智能技术的发展速度。

三是基于人工智能的基础性问题进行模块化治理。机制复合体是由不同目标任务组成的，实质是将人工智能整体挑战分解为可治理的具体模块。例如，在提高人工智能赋能的广泛性和公平性中增强发展中国家的话语权，在规范军事人工智能使用中界定责任和责任措施等。此外，机制复合体可以根据问题的性质、相关行为体的利益和能力以及地缘政治考虑来调整组成人员，在某些全球议题上需要依赖联合国及其机构等具有普遍成员资格的政府机构广泛参与；部分特定问题的解决则需要限定主体范围；还有一些问题则需要科技公司、公民社会与政府共同解决。

2.基于中心，依赖多主体的协同治理

人工智能全球治理机制复合体可整合和引导碎片化的治理，将原则、政策、规范和法律结合起来，既包括“自上而下”的科层化模式，也涉及“自下而上”的治理嵌构模式，形成有效互动。

① Laurin B Weissinger，“AI，Complexity，and Regulation”，The Oxford Handbook of AI Governance，May 19，2022，p.10.

② 张凌寒、于琳：《从传统治理到敏捷治理：生成式人工智能的治理范式革新》，载《电子政务》2023年第9期，第2-13页。

③ “Artificial Intelligence Roadmap 2024”，Department of Homeland Security，March 18，2024，https://www.dhs.gov/sites/default/files/2024- 03/24_0315_ocio_roadmap_artificialintelligence- ciov3-signed-508.pdf.

④ 陈钟、谢安明：《人工智能安全挑战及治理研究》，载《中国信息安全》2023年第5期，第32-35页。

首先，联合国应在机制复合体中扮演重要的协调角色。目前学界意识到机制复合体并不是“等级缺失（Lack of Hierarchy）”，而是具有一定等级结构的非正式等级机制，[①] 即机制复合体内部仍然存在核心机制和边缘机制的区分。基于既有的机制成效以及合法性，联合国在人工智能全球治理中应协调跨领域的合作，充当“伞”的角色来提供基础框架。在联合国已有工作基础上，可设立评估、透明中心作为机制复合体的核心机制。评估中心主要是分享与人工智能相关的最新科技突破，为政策制定者和公众提供一个共同的理解基线，增强对人工智能的总体认识，以建立主要国家间的共识；透明中心主要是设立国际信息交换中心，鼓励人工智能行为体承诺对人工智能系统进行透明和负责任的信息披露，包括要求各国报告其国内人工智能法规，私营企业提供有关先进民用人工智能研发的最新信息。[②]

其次，人工智能全球治理机制复合体是一个多维度的治理体系，它不是通过任何单一国家集团的紧密协调行动而出现的，而是通过不同行为体的累积努力来构建的，旨在解决复杂的全球性挑战的不同方面。一方面，增强广大发展中国家在人工智能全球治理中的参与度和话语权，推动形成普遍参与的国际机制和具有广泛共识的治理框架，弥合智能鸿沟，推进人工智能全球治理的包容性发展；另一方面，充分发挥私营部门和民间社会组织在议程设定、谈判、实施、监测、执行或评估相关标准和法规的过程中的作用，利用较为灵活的自律监管，改变规则普遍落后于技术创新的状态。

3. 基于信任，构建大国协调的责任治理

人工智能地缘政治风险阻碍了国际合作，但同时又需要在国际层面进行统筹调配，包括处理概念界定不一、资源管控安全化、伦理价值判断不同、国际标准博弈、南北鸿沟等问题，在此过程中大国作用尤为重要。减少大国之间的误判和对峙是构建人工智能全球治理机制复合体的关键。

首先，建立科学的认识和共同的认知。定义人工智能已经成为有效监

① 王明国、朱星宇：《国际制度复合体的结构及其应对策略》，载《世界经济与政治》2023年第11期，第89-163页。

② Emma Klein and Stewart Patric，“Envisioning a Global Regime Complex to Govern Artificial Intelligence”，Carnegie Endowment for International Peace，March，2024，https://carnegieendowment.org/2024/03/21/envisioning-global-regime-complex-to-govern-artificial-intelligence-pub-92022.

管的前提。如果不能准确界定适当的范围，决策者就无法知道人工智能的风险边界，进而引发一系列不确定的后果。因此，一个有效的人工智能治理体系需要包含能够提高对不同应用背景下风险集体理解的机制。这就需要各行为体加强政策沟通、知识共享、技术合作、风险管理和人员交流等，以解决人工智能关键术语、分类方法和标准不一的问题。可借鉴军备控制中的信任措施，加强沟通与协调，共享测试和评估标准，建立共通的衡量人工智能可信度的指标和方法、风险管理方法和相关工具的共享知识库，共享最佳实践，形成共同的战略概念话语。

其次，制定互操作性的标准和协调性法规。标准是实现人工智能底层技术互操作性的关键，推动使用兼容的标准和定义是解决当前人工智能治理碎片化的关键。一方面，可继续推动包括国际电工委员会（IEC）等组织在推进各国人工智能设计过程中的一致性；另一方面，可在国际上建立一个专门的人工智能标准组织，联合来自公共部门、行业和学术组织的不同利益相关者群体，大力促进特定领域人工智能标准的国际对话，加大基础标准制定的协商和对话，促进标准之间的互操作性。

第三，提供公共产品实现人工智能赋能的平等性和广泛性。人工智能机制复合体应确保尽可能多的国家和公民都能从人工智能中受益。这就需要在人工智能发展方面处于领先地位的国家积极提供公共产品，向低收入和中等收入国家分配关键的人工智能投入，如算力支持、构建新模型所需的硬件和软件、数据和人员培训等。

第四，加强信息共享和危机管控。大国之间需要达成广泛的共识以共同阻止人工智能的恶意使用，减少人工智能军备竞赛等。由于当前军事人工智能的国际谈判陷入僵局，难以就集体安全措施达成广泛的多边协议，对此，主要大国可以从范围较窄、不太正式的信任措施和行为准则开始，促进信息共享，并尝试建立联合预警机制等。

六、结语

基于人工智能内生技术安全风险、衍生应用安全风险，以及国家安全风险和地缘政治风险，形成了多样化的治理路径和制度，并呈现出重叠、冲突等机制复杂性的特征。当前国际上主导的治理机制、格局以及理念不

能有效应对人工智能安全威胁的激增，集中体现在治理机制未能反映人工智能安全风险特点，治理格局未能反映人工智能行为体力量变化以及互动关系，治理理念未能反映人工智能技术发展趋势。

机制复合体作为当前全球治理的重要路径，与人工智能全球治理体系较为契合。从理论框架来看，机制复合体的多中心、多模块框架与人工智能全球治理碎片化特点适配性高；从功能角度来看，机制复合体的有机协作可有效应对人工智能全球治理困境；从现有治理态势来看，人工智能机制发展趋势体现了机制复合体的特征。人工智能全球治理机制复合体需要针对人工智能安全风险特点实行基于分层、模块化的动态治理；基于中心，建立多维度的协同治理体系；促进大国之间的信任，构建大国协调的责任治理。

全球人工智能治理：多元化进程与竞争性图景

封　帅　薛世锟　马依若

摘　要：人工智能技术的快速发展对国际体系产生了重要影响。国家维度可以成为分析和展示当前全球人工智能治理进程的重要视角。从国家维度展开全球人工智能治理全景图，将会呈现出以三条核心路径为基本架构的多元化进程，分别是美国作为唯一代表的权力目标导向路径，以维护智能领域的霸权为基本特征；以欧盟等为代表的安全导向路径，其突出特征是监管优先；以中日等国为代表的发展导向路径，将治理工具与发展战略结合，希望实现技术和产业的赶超。在无政府状态的制约下，多元化与竞争性将成为国家维度下全球人工智能治理的主要特征。

关键词：人工智能治理；全球图景；国家维度；多元化进程；竞争性图景

作者简介：封帅，上海国际问题研究院国际战略与安全研究所副研究员；
薛世锟，上海国际问题研究院硕士研究生；
马依若，上海国际问题研究院硕士研究生。

一、引论

2023-2024年是人工智能发展史上的重要时刻，以ChatGPT为代表的生成式人工智能在技术创新与产业发展方面取得了突破性进展，不仅强势延续了人工智能技术在资本市场上的想象空间，而且令所有主权国家都不得不正视其在政治、军事等高政治领域的巨大潜能。时至今日，即使是最保守的研究者也不得不承认，人工智能将对国际体系变迁甚至人类社会的基本形态产生深刻影响。

感谢匿名评审专家和编辑部对本文提出的意见和建议，文责自负。

为了避免人工智能爆发式增长所带来的变革冲击人类自身的安全，对其发展方向与进程进行合理的规范便在某种意义上成为人类的共同利益。[①] 于是围绕着人工智能治理的探讨如同雨后春笋般兴起。然而，由于人工智能治理议题本身所具有的多维度特征，源于不同立场、不同意图与不同视角的研究成果在短时间内爆发式增长，使得全面而清晰地理解人工智能治理进程变得非常困难。处于不同维度的各种讨论议题交杂在一起，不仅容易出现大量的重复性研究，而且无形中造成了更高的社会传播壁垒，影响管理部门对于前沿治理理念的了解和接受度。有鉴于此，笔者选择从国家维度入手，尝试梳理当前全球人工智能治理进程的总体形态与特征，并希望通过持续的研究逐步勾勒出人工智能治理的全景图，从而为相关研究工作和政策制定提供参考。

在过去7-8年中，人工智能治理研究问题得到了世界各国研究者较为充分的讨论。研究者围绕人工智能治理政策、路径与思想进行了各种形式的探索，取得了丰富的成果。[②] 几乎在同一时间，中国学者也积极投身于人工智能治理研究，形成了大量具有启发性的研究成果，既有关于人工智能治理规律的讨论，又有对世界各国治理政策的详细分析。[③] 这些研究成

① 参见：《全球人工智能治理倡议》，中央网信办网站，2023年10月18日，https://www.cac.gov.cn/2023-10/18/c_1699291032884978.htm。

② 参见：U. Gasser and V.A.F. Almeida, "A layered model for AI governance", *IEEE Internet Computing*, Vol. 21 No. 6, 2017, pp. 58-62; Nathalie A. Smuha, "Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea", *Philosophy & Technology*, Vol.34, Issue 1, 2021, pp.91-104; Dexe Jacob and Franke Ulrik, "Nordic lights? National AI policies for doing well by doing good", *Journal of Cyber Policy*, Vol.5, Issue 3, 2020, pp.332-349; Stefan Larsson, "On the Governance of Artificial Intelligence through Ethics Guidelines", *Asian Journal of Law and Society*, Vol.7, Issue 3, 2020, pp.437-451; M. Mäntymäki, M. Minkkinen, T. Birkstedt, and M. Viljanen, "Defining organizational AI governance", *AI and Ethics*, Vol.2, No.4, 2022, pp.603-609; Lu Qinghua etc, "Responsible AI Patern Catalogue: A Collection of Best Practices for AI Governance and Engineering", *ACM Computing Suveys*, 2023, https://dl.acm.org/doi/pdf/10.1145/3626234; Teemu Birkstedt, "AI governance: themes, knowledge gaps and future agendas", *Internet Research*. Vol.33, Issue 7, 2023, pp. 133-167. 等等。

③ 参见：杨晓雷：《人工智能治理研究》，北京大学出版社2022年版；清华大学战略与安全研究中心主编：《人工智能与治理》，中国社会科学出版社2022年版；薛澜、梁正、张辉、曾雄：《人工智能治理框架与实施路径》，中国科学技术出版社2024年版；贾开、蒋余浩：《人工智能治理的三个基本问题：技术逻辑、风险挑战与公共政策选择》，载《中国行政管理》2017年第10期，第40-45页。傅莹：《人工智能的治理和国际机制的关键要素》，载《人民论坛》2020年第4期，第6-8页；高奇琦：《全球善智与全球合智：人工智能全球治理的未来》，载《世界经济与政治》2019年第7期，第24-48页；宋黎磊、戴淑婷：《中美欧人工智能治理领域的竞争与合作》，载《当代中国与世界》2021年第4期，第58-67页；鲁传颖：《人工智能：一项战略性技术的应用与治理》，载《人民论坛》2024年第1期，第72-75页等等。

果为本文的研究提供了有益的参考。此外，世界各国围绕人工智能治理出台了大量法律法规及规范性文件，也成为本文写作的重要基础。

二、主权国家参与人工智能治理的基础逻辑

人工智能治理并不仅仅是一种国家行为，但主权国家在当前全球人工智能治理进程中毫无疑问占据着核心位置。因此国家维度是当前所有试图勾勒全球人工智能治理总体态势的尝试中最为自然的选择，也是国际关系和区域国别学学科涉足人工智能治理研究的主要立足点。

作为现代社会的核心行为体，主权国家在过去数百年的人类历史中一直牢牢占据着国际体系的核心位置，在无政府状态下已经稳定地形成了以国家利益和权力分配为核心的运行规律。在可预见的时间段里，主权国家的特征仍然是一个常量，而人工智能技术的爆发式成长则是21世纪国际体系中的突发新变量，因此，从主权国家的视角来看，人工智能技术的爆发式成长无疑是前沿技术进步所带来的新一轮挑战。因为这一挑战具有驱动系统性变革的潜力，令技术社群和社会科学研究者产生系统变革的遐想。[①] 但对于主权国家这样一个理性行为体而言，它的所有行为和政策仍需以其自身的基本运行规律为基础，思维方式的路径依赖也很难在短时间内被打破。于是，在传统的国家利益、权力规则与人工智能技术发展规律的持续纠缠与相互作用下，国家维度的人工智能治理进程展现出了复杂的特征与形态。

（一）主权国家参与人工智能治理的必然性

作为当代国际体系的主要行为体，主权国家不可能回避人工智能技术跃迁这一重大社会变化。国家行为体自然会关注人工智能社会系统的全部核心环节，但治理是其对人工智能系统运转进行干预的主要抓手，因此，主权国家深度参与人工智能治理进程具有逻辑上的必然性。

如图1所示，人工智能技术发展所带来的巨大的现实影响力源自其所带来的系统性进程，其中最为关键的支撑性环节当属技术创新、产业发展与治理体系建设。[②] 三者相互联系、相互影响，形成了一个完整的系统。

① 笔者对于可能的未来变化图景也有过相应的设想与讨论，参见：封帅：《从民族国家到全球秩序：人工智能时代的世界政治图景》，载《外交评论》2020年第6期，第99–129页。

② 参见：杨晓雷：《人工智能治理研究》，北京大学出版社2022年版。

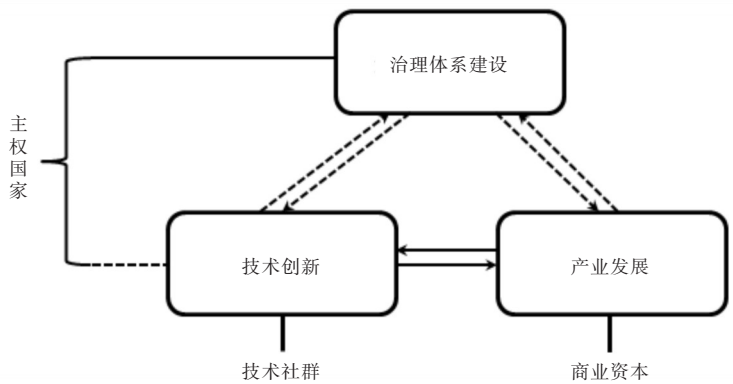

图1：人工智能系统的关键环节与多元主体参与路径示意图

资料来源：作者自制

在该系统中，技术创新与产业发展相互促进，形成了紧密的纽带联系。而治理体系建设则尝试对技术创新和产业发展形成规制，防范各种安全风险，并对系统运行加以引导。

基于人工智能技术进步所形成的庞大系统是复杂的，任何性质的行为主体要想在这个复杂系统中获得发言权，就必须有能力影响其中某个关键环节的运行。技术社群的影响力来源于其在技术创新环节不可替代的作用，商业资本则通过有效驱动智能产业的发展而获得影响力。而作为现代国际体系中的主导力量，主权国家虽然在理论上有权力参与该系统的所有环节，但在现实中，人工智能技术特点与智能产业发展规律都为这两个环节设置了较高的专业壁垒，主权国家并无太多发挥作用的空间。只有治理体系建设环节与其自身的功能天然契合，自然成为主权国家参与人工智能事务、发挥自身影响力的主要抓手。

事实上，在当代世界体系中，所有的重要治理行为都无法绕过主权国家而独立存在。一方面，在主权国家内部，所有针对人工智能的治理内容最终都需要以国家法律、法规、规章等形式予以确认，才能够获得广泛的效能。主权国家是各种治理资源和治理思想得以凝结并转化为治理能力的关键行为体。另一方面，主权国家是建立和推广国际规则的核心主体。在国际无政府状态下，任何治理的国际规则都不可能在未经主权国家授权的情况下获得普遍的贯彻与推广，缺少主权国家——特别

是关键大国——的共识与协调，国际规范与规则难以形成。主权国家因其特征与人工智能治理议题形成了共鸣，迅速取代其他行为体，成为人工智能治理进程的主导者，并主要通过治理体系建设的方式来影响和塑造人工智能社会系统的发展。[①] 这也就意味着，在描绘当前全球人工智能治理进程时，国家维度是一个能够充分展示该进程全貌的合理选择。[②]

（二）主权国家推动人工智能治理：理想与现实

虽然主权国家希望通过自己的治理方案有效塑造人工智能系统，但从现实情况来看，主权国家针对人工智能治理的完美设想在实践中很难实现，人工智能技术规律与国家治理规律之间的张力决定了人工智能治理进程很可能出现复杂的非平衡状态。

对于主权国家而言，开展人工智能治理仍然离不开传统的国家治理经验，因此，几乎所有主权国家在参与人工智能治理活动时，都会希望达到这样一种理想的状态：即在各方利益充分表达的基础上，经过广泛的社会动员，形成一个有效且稳定的治理体系。在该体系的规制下，人工智能技术能够在安全可靠的环境中持续创新，并凝结为产业发展成果，进而推动人工智能技术在军事安全以及其他高政治领域得到有效影响，确保国家利益的实现。这种理想模型结构如图2所示，这种理想模型的希冀在于，安全、稳定与发展将在该模型中同时实现，并且令人工智能系统各主要环节达到平衡状态。[③]

然而，无论是从现实案例还是理论分析来看，这种人工智能治理的理想模型都难以在现实世界中得到落实。最为关键的原因在于治理体系与技术人工智能技术的变化速度存在巨大落差。

作为一项潜在的颠覆性技术，复杂性、突变性与不确定性一直是人

① 以美国2016年《国家人工智能研究与发展战略规划》的颁布为标志，主权国家开始实际介入人工智能治理进程，并迅速成为人工智能治理议程的主导力量，从2016至2022年，与“人工智能”相关的法案数量从1项增长到37项，充分展现了主权国家对于人工智能治理议题的高度重视。参见：“Artificial Intelligence Index Report 2023”，HAI，April 2023，p.268，https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf。

② 相关讨论还可参见：薛澜、梁正、张辉、曾雄：《人工智能治理框架与实施路径》，中国科学技术出版社2024年版。

③ 参见：Araz Taeihagh，“Governance on Artificial Intelligence”，*Policy and Society*，Vol.40，No.2，2021，pp.137-157。

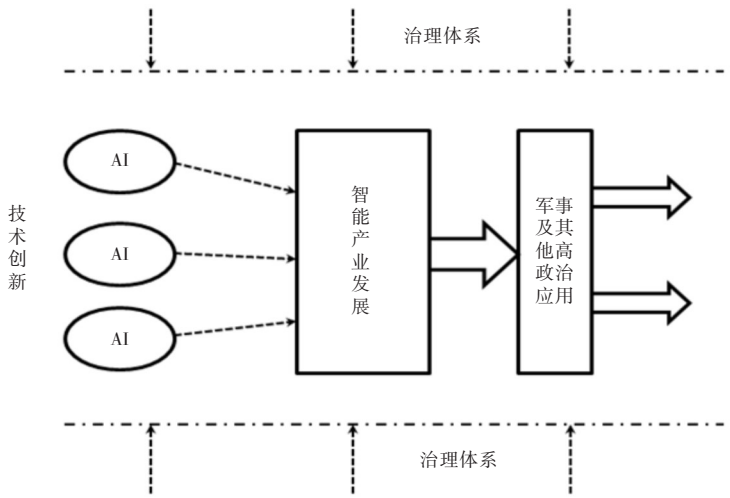

图2：主权国家设想的人工智能治理理想模型

资料来源：作者自制

工智能的重要特点。其突破性的技术变革出现后，往往会以加速的方式向前推进，在短时间内取得重大进展。人工智能技术与社会系统的融合过程也非常复杂，其所产生的社会和安全影响常常难以进行提前评估。更为重要的是，人类对于人工智能的了解尚待深入，对于其未来发展还不能够形成非常确定的评估。对于人工智能技术的诸多焦虑也正源于这种高速率的变革。但是，相对而言，主权国家建构治理体系的节奏却要慢很多，作为一种国家行为，传统治理活动有着严格的流程与特征，从安全风险认知到各方利益表达，从社会动员到社会共识形成，从法案设计、讨论、制定到审议和通过，再到机制设立与运行，将会经历很长的一段时间，而且在现有的政治架构中，这些环节也是不可或缺的。在不同的运行节奏下，主权国家的人工智能治理现实场景常常会在以下两种状态中摇摆。

第一种情形如图3所示，该模型表现的是一旦治理规范超前于技术发展进程，虽然通过较为严格的限制控制了技术对于国家与社会的冲击，但治理体系对于技术创新和产业发展形成了限制和挤压，将失去部分技术创新空间，阻碍新兴产业发展。

第二种情形如图4所示，该模型表现的是当治理活动严重滞后于技术

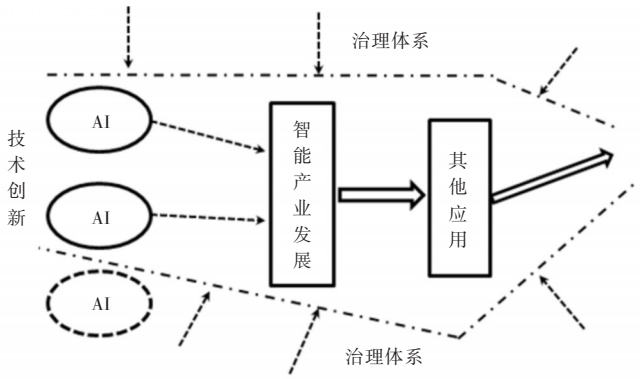

图3：人工智能治理现实模型（Ⅰ）：治理超前

资料来源：作者自制

与产业发展的需要时，技术的创新实际已经超出了治理议程的设计，在各种力量的冲击下，治理进程呈现出碎片化的状态。虽然看似产业得到了充分发展，但由于失去有效的边界限制，其所累积的风险将对整个社会乃至于国际体系形成严重且难以应对的威胁。

综上，人工智能技术的跨越式发展将会使主权国家面临前所未有的考验，稳健的传统治理模式与高速变革的技术发展节奏将使建立长期稳定治理体系的尝试变得异常艰难。在理论上，面对快速变革的治理对象，只有保持足够的“灵活”与“敏捷”，才能够暂时维护人工智能系统各环节的平衡运

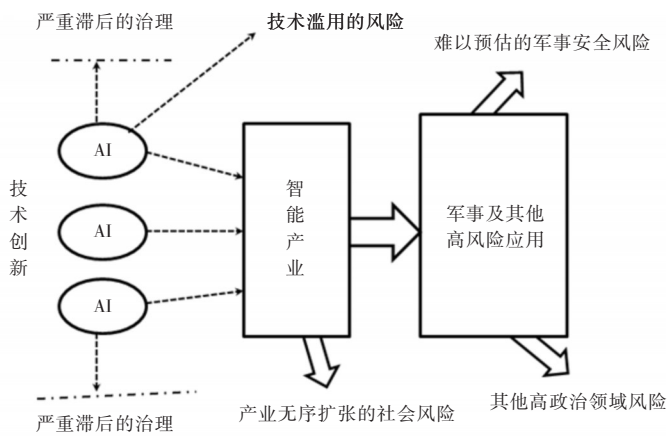

图4：人工智能治理现实模型（Ⅱ）：治理滞后

资料来源：作者自制

转，这对传统治理体系提出了重大的变革需求。[①] 然而，主权国家的基本特征与政治运行方式在短时间内难以出现重大改革。于是，我们在政治现实中观察到的现象主要是主权国家不得不在超前和滞后这两种治理模型中做出选择，而最终呈现出来的那种复杂状态，恰恰是不同选择的结果。

三、国家维度的路径选择：安全、发展与权力

在主权国家视野中，国家利益是一个综合性的框架，人工智能在多个方面的影响都能够在国家利益框架内产生回响。所以主权国家在参与人工智能治理进程时，就必须通盘考虑各方面的利益影响，结合国家战略目标，对于不同方面的内容进行优先级排序，在有所取舍的情况下创建具有各自特色的治理体系。理论上来说，可选择的方案有很多种，但从实践上看，受到各国技术基础、市场规模、文化传统等客观条件的制约，国家在人工智能治理方面选择的空间又是相对有限的，目前较为合理的路径大体上只有三条。

第一，安全目标导向的治理路径。安全是所有治理行为的基本目标，通过合理的治理体系建设，确保人工智能技术和产业的发展不会对国家安全造成严重的负面影响，这是所有国家开展治理活动的基本逻辑。由于人工智能技术广泛的适应性与渗透性，人工智能治理的安全目标指向多个层次。其一，避免人工智能的发展对人本身造成安全威胁。包括技术发展过程中不能对人的生命、伦理、尊严以及隐私等造成严重威胁，维护人的安全是最基础的治理目标。其二，避免人工智能的发展对国家本身造成安全威胁，包括保障不同政治体系国家的政治安全，避免人工智能武器对于国家存续造成严重威胁等。其三，减少人工智能与多领域结构深度结合造成的系统性不稳定。简而言之，就是对人工智能在经济、社会等领域可能带来的“创造性破坏”过程进行有效引导，尽量减少结构变革对于社会稳定的不利影响。[②]

① 参见：清华大学人工智能国际治理研究院、清华大学中国科技政策研究中心、人工智能治理研究中心编著：《敏捷与协同：人工治理理念与实践前沿》，社会科学文献出版社2024年版。

② 参见：Greg Allen and Taniel Chan，“Artificial Intelligence and National Security”，Intelligence Advanced Research Projects Activity（IARPA），Belfer Center，Harvard Kennedy School，July 2017；封帅：《国家安全学视阈下的人工智能安全研究：议题网络建构的初步尝试》，载《国际安全研究》2023年第1期，第26–49页。

第二，发展目标导向的治理路径。对于绝大部分主权国家来说，人工智能技术的发展都同时意味着机遇和挑战，挑战来自人工智能技术所带来的安全风险，而机遇则来自于人工智能所孕育的巨大发展潜能。在人工智能技术变革浪潮一轮又一轮的冲击下，各国已经普遍接受这一判断，即人工智能技术将成为未来经济发展的核心驱动力，新经济业态的创造者，也是新一轮工业革命的关键力量。这些内容都是国家利益的核心组成部分，是国家实力的根源。因此，在主权国家的行为逻辑中，为防范安全风险，以治理体系建设的方式对人工智能进行一定程度的限制只是必不可少的手段，而非唯一的目的。治理的另一重要目标是确保人工智能技术得以快速发展和充分利用，并有效促进经济社会发展，提升国家的综合实力，因此，在发展与安全之间寻找动态平衡点，就将成为国家维度人工智能治理的重要任务。

第三，权力目标导向的治理路径。在国际无政府状态下，权力目标仍将是人工智能治理进程中国家行为的深层动力，这也是主权国家与其他类别行为体之间最明显的区别。国际关系研究离不开权力的概念，权力是对主体间互动关系的一种界定和表述，“A 对 B 拥有权力，即 A 能让 B 做某些 B 不愿做的事情”。[①] 该路径的实质是在处理人工智能问题时，始终以保持在该领域的权力优势为基本目标。因此，需要同时在人工智能系统的三个关键方面都保持充分的竞争力。这种路径对于行为体能力的要求非常高，但对于人工智能技术的领先国家而言，该路径的吸引力不言而喻。

从国家维度展开观察，当前全球主要大国所选择的人工智能治理路径大体上都沿着上述路径展开。留给世界各国开展人工智能治理路径选择的时间窗口是非常有限的，一旦进入了相应的发展路径，往往会在较短时间内形成相应的路径依赖。我们将根据此框架对当前全球人工智能治理领域较为活跃的主权国家以及作为整体的欧盟[②] 的治理活动进行分析和概括，

① Robert Dahl, “The Concept of Power,” *Behavioral Science*, Vol.2 No.3, 1957, p.203.

② 需要在此特别说明的是，本文的案例均选自主权国家人工智能治理活动的相关内容，唯一的例外是将欧盟作为与主权国家并列的独立个体进行研究。这是由于欧盟作为目前全球最为成功的政治一体化组织，在人工智能治理领域具有非常明显的一体化特征，对其成员国影响最大的人工智能治理活动都是在欧盟层面上推进的，而且已经形成了相应路径的典型代表。可以说在人工智能治理领域，欧盟已经成为一个结构紧密、具有与主权国家类似特征的行为体。欧盟在该领域的治理政策与方案是其成员国共同商讨的结果，也得到所有成员国的尊重与执行。因此，笔者根据本领域研究的惯例将欧盟作为独立的行为体与主权国家放在同一层次中，这种特殊安排也是尊重现实情况的表现。目前尚无其他地区一体化组织能够达到这种程度，但不排除随着人类社会的发展，未来还会有类似机构以同样的逻辑加入国家维度的讨论。特此说明。

以其自身的资源禀赋、发展阶段与国家战略等基本状态为“经”，以其在面对人工智能技术跃迁时对于安全目标、发展目标和权力目标的不同排序为“纬”，展示出当前全球人工智能治理进程中广泛存在的多种治理模式平行推进且相互竞争的基本图景。

四、全球人工智能治理：选择与竞争塑造的多元进程

在短短几年的时间里，主权国家在人工智能治理方面取得了重要的成果，形成了数量庞大的原则、法案与文本。但迄今为止，主权国家并未在人工智能治理议题上形成具有压倒性优势的主流路径。不同国家在开展治理活动时，都会综合考量多方要素，结合自身的历史传统、战略思路、思想理念等，在目标体系中作出不同的优先级排序，进而导入不同的治理路径。在选择与竞争中造就的多元进程构成了全球人工智能治理的基本形态。

（一）权力目标导向的人工智能治理：美国维护智能霸权的野心

在当前国际体系中，只有美国较为明确地选择了以权力目标为导向的人工智能治理路径，目前也几乎只有美国能够做出这样的选择。

作为当代国际体系中唯一的霸权国，美国在人工智能技术发展史上一直扮演着关键角色。在本轮人工智能技术浪潮中，美国仍然保持着明显的领先地位。按照当前全球人工智能发展水平的一般评估标准来说，无论从论文、人才、投资、产业等任何一个方面进行同口径国别比较，美国都是当仁不让的领先者之一。正是因为具备如此雄厚的基础，美国政府在应对人工智能挑战的时候，选择了权力目标导向路径，即同时推进技术、产业和治理目标的实现，维持并扩大美国在人工智能领域的领先优势，将人工智能作为维护霸权的有效武器。

如表1所示，从2016年至今，美国已发布20多份与人工智能治理直接相关的战略文件，已经基本形成了系统的治理体系。当前对美国人工智能治理机制影响最大的法案当属《2020年国家人工智能计划法案》以及根据该法案实施的“国家人工智能计划”。[①] 根据该法案的要求，美国建

① 116th US Congress，H.R.6216-National Artificial Intelligence Initiative Act of 2020，March 12，2020，https://www.congress.gov/bill/116th-congress/house-bill/6216？__cf_chl_tk=aq2aBo0eMP-PRwKVkU0EWnlBuhMiHEul5V90n1R.mpY8-1653477655-0-gaNycGzNBz0.

表1：美国人工智能治理相关文件列表

时间		颁布机构	文件名称
奥巴马政府	2016.10	白宫/国家科学技术委员会	国家人工智能研究与发展战略规划
	2016.10	白宫/国家科学技术委员会	为人工智能的未来做好准备
	2016.12	总统行政办公室	人工智能、自动化与经济
特朗普政府	2019.2	白宫	保持美国在人工智能领域的领先地位
	2019.6	白宫/国家科学技术委员会	国家人工智能研发战略计划：2019年版
	2019.10	国防创新委员会	人工智能原则：国防部应用人工智能伦理建议
	2019.11	白宫科技政策办公室	2016–2019年人工智能研发进展
	2020.2	白宫	国家人工智能倡议：首个年度报告
	2020.2	白宫	促进在联邦政府中使用可信赖人工智能
	2020.3	美国国会	2020年国家人工智能计划法案
	2020.8	美国国家标准技术研究所	可解释的人工智能四项原则
	2020.10	白宫	关键和新兴技术国家战略
	2020.12	白宫	促进在联邦政府中使用可信赖人工智能
拜登政府	2021.1	美国国会	2020年国家人工智能计划法案
	2021.3	人工智能国家安全委员会	委员会最终报告（2021年版）
	2021.8	国土安全部	人工智能/机器学习战略计划
	2022.2	美国国家标准技术研究所	关键和新兴技术国家战略
	2022.3	美国国家标准技术研究所	迈向识别和管理人工智能偏见的标准
	2022.10	白宫科技政策办公室	人工智能权利法案蓝图：让自动化系统为美国服务
	2023.1	美国国家标准与技术研究	人工智能风险管理框架
	2023.5	白宫国家科学技术委员会	国家人工智能研发战略计划：2023年版
	2023.10	白宫	关于安全、可靠、可信地开发和使用人工智能的行政令

资料来源：作者自制

立了以国家人工智能计划办公室（NAIIO）、国家人工智能咨询委员会（NAICA）、人工智能专责委员会（SCAI）、国家人工智能研究资源工作组（NAIRRTF）为主要代表的治理机制，该机制建设具有系统性。[②] 当然，人工智能治理本身是一个影响广泛的议程，其国内各种主体的行为都会与此产生牵涉。

美国在人工智能治理领域的思路具有非常明显的独特性，意图维持超

② 该机构的相关情况可参见：The National Artificial Intelligence Initiative Office，https://www.ai.gov/naiio/；The National Artificial Intelligence Research Resource Task Force，https://www.ai.gov/nairrtf/等等。

级大国的思想意蕴非常清晰。

首先，从根本上说，美国参与人工智能治理的目标是寻求高效发展与强大的国际竞争力。美国政府在国家人工智能计划法案及多个文件中都将研究和开发置于首位，强调要大力投资人工智能技术的研发，并将了解人工智能理论能力和局限性、促进类人人工智能研究等多项研发领域作为优先项。[①] 2019年2月11日，时任美国总统特朗普签署《保持美国在人工智能领域的领先地位》（Maintaining American Leadership in Artificial Intelligence）的行政令。[②] 此后的各种文件共总结出六种为保持美国在全球人工智能领域领先地位的途径，包括投资人工智能研发、开放人工智能资源、消除人工智能创新的障碍、培养符合人工智能领域需求的劳动力、促进形成支撑美国人工智能创新的国际环境以及为实现政府服务和使命采纳可信赖的人工智能技术。

其次，美国的人工智能治理方案多为“软法”，常常采取总统行政令的方式发布。落实过程主要从两个路径切入：一是联邦政府多部门各自发布人工智能应用与治理战略，二是市场与企业“自产自销”治理规则。美国联邦政府多个部门率先发布人工智能发展战略与实际应用的自我约束机制，例如2020年12月《美国国土安全部人工智能战略》、2020年12月《美国司法部人工智能战略》、2021年1月《美国卫生与公众服务部人工智能战略》、2021年7月《退伍军人事务部人工智能战略》，这些战略使得美国率先在官方部门形成了较为完整的人工智能治理体系。[③] 同时，在企业内部与市场范围内，美国政府采取不过多干预政策，在符合治理价值观的基础下放任人工智能产业发展和其他企业运用人工智能技术。美国政府在其中主要起到“盲区补位”的作用，即“要重点关注有重大社会效应但产业不太可能涉及的领域”，如“用于公共卫生、城市系统和智能社区的AI

① “The National AI Research and Development Strategic Plan：2019 Update”，National Science and Technology Council，Networking and Information Technology Research and Development Subcommittee，June 2019，https：//www.Nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf.

② “Executive Order on Maintaining American Leadership in Artificial Intelligence”，The White House，February 2019，https：//www.Whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/.

③ “Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government”，The White House，December 3，2020，https：//trumpwhitehouse.archives.gov/presidential-actions/executive-order-promoting-use-trustworthy-artificial-intelligence-federal-government/.

领域，以及涉及社会福利、刑事司法、环境可持续发展和国家安全相关的领域”。[①]

最后，美国具有侧重战略安全与“去偏见”的治理偏好。在国内安全方面，美国强调人工智能的发展不能破坏美国的社会稳定与民主价值观，更不能侵害美国民众的个人隐私等权利，要求在治理的过程中制定人工智能安全指南、标准和最佳时间，保障技术安全。[②]受传统的社会价值观与美国民主党特色的影响，美国政府强调人工智能要非歧视性发展，防范系统性、统计性和人类性偏见，保证数据集的代表性和算法的公平性。[③]但在国际安全方面，美国所注重的则是一种对技术相对权力最大化的追求。美国人工智能国家安全委员会提出要整合国家战略、调整国家定位、团结盟友和合作伙伴来应对人工智能时代的冲突和竞争，以期实现在人工智能时代保卫美国和赢得技术竞争。为实现这一目的，美国在国际人工智能领域积极推动数字地缘政治竞争，不惜采取非市场行为的恶性竞争也要限制竞争对手的发展。[④]

总之，作为人工智能领域的超级大国，美国是唯一选择权力导向路径参与人工智能治理的国家，美国的人工智能治理方案与治理路径具有系统性与完整性，覆盖政府、社会、市场、国内与国际全范围，很多具体安排都可资借鉴。但从全球层面来看，美国的人工智能治理行为非常明确地显露了维持美国智能霸权的野心，意图在国际人工智能竞争中取胜借以维持国际领导地位，与全球化时代的需求背道而驰，在很大程度上也成为全球人工智能领域恶性竞争的源头之一。

---

① “The National AI Research and Development Strategic Plan：2019 Update”，National Science and Technology Council，Networking and Information Technology Research and Development Subcommittee，June 2019，https：//www.Nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf.

② “Executive Order on the Safe，Secure，and Trustworthy Development and Use of Artificial Intelligence”，The White House，October 30，2023，https：//www.whitehouse.gov/briefing-room/presidential- actions/2023/10/30/executive- order- on- the- safe- secure- and- trustworthy- development- and-use-of-artificial-intelligence/.

③ “Towards a Standard for Identifying and Managing Bias in Artificial Intelligence”，NIST，March 15，2022，https：//www.nist.gov/publications/towards-standard-identifying-and-managing-bias-artificial-intelligence.

④ 相关情况可参见：鲁传颖：《全球数字地缘政治的战略态势及其影响》，载《当代世界》2023年第5期，第37-43页。

（二）安全目标导向的人工智能治理：以监管争夺话语权

除美国之外的全球其他国家在面对人工智能技术浪潮时，都无法保持系统性的领先，所以他们需要在技术、产业和治理三个方面做出排序和取舍。以欧盟、英国和印度等为代表的部分行为体基于多重考虑，选择了安全目标导向的人工智能治理路径，即聚焦监管，并尽快形成系统的治理体系，以此作为获得人工智能领域话语权的支点。

欧盟是安全目标导向路径的主要代表，它在人工智能治理方面的建设也具有典型性。欧盟选择以监管优先的方式具有相当程度的必然性，一方面，该路径与欧盟在数字治理方面的整体思路一脉相承，是数据流通治理体系的自然延续。在人工智能技术得到关注之前，欧盟对于网络空间的数据流通等问题就采取了强监管的措施。这种强监管思路不仅与欧洲国家关注个人隐私，严密防止入侵个人私域的社会文化深度契合，也为国家行为体介入数字空间事务提供了有效依托。有这样的经验积累，在面对人工智能技术时延续强监管思路顺理成章。另一方面，在技术和产业暂时落后于中美等国的情况下，加强治理研究也成为欧盟获取人工智能领域比较优势与权力的一种合理选择。以2016年通过、2018年正式生效的《通用数据保护条例》（GDPR）为代表，欧盟在一定程度上引领了全球数字治理的基本逻辑。该条例已经成为全球数据治理的经典文献，对全球数字治理的整体方向产生了重要影响。欧盟希望在人工智能领域继续复制这样的经验，而承载这种希冀的就是欧盟《人工智能法案》（AI ACT）。

如表2所示，虽然欧盟颁布的文本数量少于美国，但从法案内容上看，欧盟在人工智能治理领域的严格程度远超其他所有国家。欧盟对人工智能的高速发展表现出强烈的焦虑，它对于智能技术对人类社会和人本身可能产生的侵蚀非常担忧。为了防范所有指向人的风险，欧盟意图构造系统、全面的监管人工智能的硬规制体系。

欧盟虽然在人工智能发展领域也有所规划，但在政策上显然更加重视讨论人工智能对法律与道德的挑战。欧盟强调发展与应用的人工智能一定是“可信赖的”。“可信赖的”人工智能应该合乎且尊重伦理原则和价值观，从技术与社会环境角度确保人工智能的稳健发展。[①] 只有这样才能防

① “Ethics guidelines for trustworthy AI”，European Commission，April 8，2019，https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

表2：欧盟人工智能治理相关文件列表

时间	颁布机构	文件名称
2018.4	欧盟成员国签署	人工智能合作宣言
2018.4	欧盟委员会	欧盟人工智能战略（Artificial Intelligence for Europe）
2018.12	欧盟委员会	人工智能协调计划
2019.4	人工智能高级别专家组	可信赖人工智能伦理指南
2019.4	人工智能高级别专家组	可信赖人工智能的政策和投资建议
2020.2	欧盟委员会	人工智能白皮书：欧洲实现卓越与信任的欧洲之路
2020.9	人工智能高级别专家组	可信赖人工智能评估清单
2020.6	欧盟委员会	欧洲数据战略
2021-2024	欧盟各机构共同参与	制定《人工智能法案》各项工作

资料来源：作者自制

止人工智能对个人基本权利的威胁，如个人数据和隐私保护，或偏见歧视问题，以及应对人工智能对安全风险和责任制度的挑战。

为实现这一目标，欧盟在全球范围内首先推动建立人工智能监管的综合性法律，希望通过“硬法”的创建确保欧盟范围内人工智能“可信赖”，并且能够充分体现欧盟价值观。因此，欧盟的政策偏好是确保算法透明、系统性分级管理与完善问责制。欧盟要求企业必须重点保护个人数据，提出数据的处理与企业人工智能的算法必须合法、公正和透明。为确保治理方案的合理性与有效性，欧盟采取基于风险的分级监管治理思路，将人工智能技术划分为四个等级：不可接受风险、高风险、有限风险和最小风险。对最小风险的应用监管可能只是要求报备使用记录，而对一些定义为高风险的应用可能会是以强制性要求进行监管，部分被划为不可接受的风险的人工智能系统将被直接禁止。[①] 为助力治理规则的监督和执行，欧盟要求人工智能系统必须在技术上强大且准确、高风险人工智能都需要接受人为监督、明确高风险人工智能应用的法律归属责任，将参与人工智能系统生命周期的相关者都视为主要责任主体。

2024年3月，《人工智能法案》在欧洲议会投票通过，并将于5月得到正式批准。这将是人工智能治理领域的一个标志性时刻，“硬法”体

① “Artificial Intelligence Act”，European Parliament，March 2014，https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf.

系的创建将使人工智能治理强度得到明显提升，其未来的执行效果将会对欧盟的技术和产业发展带来重大影响，也会为其他人工智能治理议程提供参考。

除欧盟之外，英国、印度等国也倾向于选择监管优先的治理思路，也可以被视为安全目标主导路径的成员。

英国是全球最早关注人工智能技术发展潜能的国家之一，在全球技术与产业竞争方面也占据一席之地。在很长时间内，英国曾被视为在人工智能领域选择发展目标导向的国家之一，其在2022年发布的《人工智能监管政策》中表现出较为宽松的治理思路。但在苏纳克（Rishi Sunak）执政之后，英国一改原有政策思路，转而高调宣传人工智能安全问题，不仅在短时间内推出各种治理文件，而且于2023年11月1日主办了首届全球人工智能安全峰会，并推动28国共同签署了《布莱奇利宣言》（Bletchley Declaration），这也是全球第一份针对人工智能技术的国际性声明。[①] 英国领导人在各种场合不断表达英国要成为人工智能治理领域主导国家的意愿，希望能够在监管方面“领导全球”。[②]

印度作为在人工智能领域具有较大潜力的发展中国家，目前在人工智能治理领域也显示出监管优先的特征。印度的人工智能治理策略明显参考了欧盟方案，整体显示为同时推进的两条治理思路，一是法律法规管理与监督人工智能系统，即在重要行业、领域制定针对性的法律法规，防止人工智能技术的负面影响。二是基于技术管理人工智能系统，即利用统计等技术手段来感知随着人工智能不断发展，治理相关的新兴热点话题。[③] 从整体趋势来看，印度对于人工智能的管控在不断加强，特别是对于能够产生广泛社会影响以及可能影响大选的社交媒体平台的限制越来越多。有消息称，印度也正在考虑在人工智能领域推动立法。[④]

① “首届人工智能安全峰会发布《布莱奇利宣言》，新华社，2023年11月3日，http://www.news.cn/world/2023-11/03/c_1129955096.htm。

② 参见：“A pro-innovation approach to AI regulation”，UK Policy Paper，August 3，2023，https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper。

③ “Responsible AI：Part 2-Operationalizing Principles for Responsible AI”，NITI Aayog，August 2021，https://www.niti.gov.in/sites/default/files/2021-08/Part2-Responsible-AI-12082021.pdf.

④ ［印］亚什拉吉·夏尔马：《印度莫迪政府在全国大选前急于监管人工智能》，半岛电视台，2024年3月13日，https://chinese.aljazeera.net/news/2024/3/13/印度莫迪政府在全国大选前急于监管人工智能。

总之，以欧盟为主要代表的国家群体选择安全目标主导路径参与人工智能治理活动。该路径主体具有对于安全风险更加敏感，更加倾向于依靠硬法来开展治理活动等特征。这一路径对于那些有较大规模数字市场，希望在全球人工智能领域争夺话语权，但在技术和产业发展方面存在瓶颈的行为体具有很大吸引力。由于对监管的高度重视，仅从治理角度来看，选择该路径国家的实践活动极大地推动了人工智能治理思想与方法的完善，包括人工智能治理写进立法议程，基于风险将人工智能进行分级管理等都是卓有成效的创举，也确实在一定时间内实现了国际影响力提升的目标。但该路径始终无法摆脱的矛盾是，过多法律规制与监管介入必然阻碍人工智能产业的研发速度与发展效率，据估算，为满足《人工智能法案》的要求，最大程度上会使欧盟人工智能企业的成本增幅达到17%，这将严重削弱企业的全球竞争力。[①] 一旦失去技术和产业方面的长远发展能力，在治理领域取得的权力优势最终也将因为缺少充分支撑而逐渐耗散。如何平衡监管与发展将是选择该路径的国家需要进一步攻克的难题。

（三）发展目标导向的人工智能治理：为跨越式发展提供基石

当一个国家在面对人工智能技术浪潮冲击，优先选择推动技术和产业发展时，大体上就意味着它选择了发展目标导向路径。该路径强调要充分利用人工智能技术推动各领域的跨越式发展，避免在新兴技术引领的国际竞争中陷入不利局面。因此，该路径往往选择将治理要求置于整体的战略发展规划之内，在保证技术和社会安全的基础上，尽量减少对技术创新和产业发展的影响。

选择该路径的国家大体上具备两方面的特点：其一是具有强烈的发展意愿，并且有意愿迎接技术变革的挑战；其二是具有一定基础条件，即国内存在契合人工智能技术发展的资源禀赋，有可能在新的技术变革背景下赢得新的发展机遇。前者是国家的主观意愿，后者是国家所具备的客观条件，选择该路径的国家实际上做出了在人工智能领域追赶前沿国家的战略选择。目前，中国、日本、俄罗斯等国可以被视为该路径的典型代表。

中国是以发展目标导向路径参与人工智能治理最为典型的国家。中国政府将以人工智能为代表的数字技术革命视为“中华民族伟大复兴的重大

① 封帅：《欧盟通过<人工智能法案>传递了什么信号？》，载《新民晚报》2024年3月16日。

表3：中国人工智能治理相关文件列表

时间	颁布机构	文件名称
2017.7	国务院	新一代人工智能发展规划
2017.12	工业与信息化部	促进新一代人工智能产业发展三年行动计划（2018-2020）
2019.8	科技部	国家新一代人工智能开放创新平台建设工作指引
2019.6	国家新一代人工智能治理专业委员会发布	新一代人工智能治理原则——发展负责任的人工智能
2020.7	中央网信办	国家新一代人工智能标准体系建设指南
2021.9	国家新一代人工智能治理专业委员会	新一代人工智能伦理规范
2023.8	中央网信办、国家发展改革委、教育部、科技部、工业和信息化部、公安部、广电总局	生成式人工智能服务管理暂行办法
2023.10	中央网信办	全球人工智能治理倡议

资料来源：作者自制

历史机遇”。[①] 如表3所示，从2017年至今，中国各部门围绕国家人工智能的发展颁布了多项规划、原则与倡议。中国在人工智能治理领域活动的一个显著特点是，治理议题的相关内容被纳入人工智能发展的总体规划框架内，作为发展规划的重要组成部分为人工智能的发展保驾护航。

在党中央的统一领导下，根据国务院颁布的《新一代人工智能发展规划》的系统规划，包括国家科技领导小组、发改委、工信部、科技部以及各科学研究管理机构等都共同参与到人工智能治理活动中，在该框架内设立的“新一代人工智能”发展规划推进办公室则成为中国人工智能治理机制运行的主要协调机构。[②] 通过合理的顶层设计与协调联动，逐渐形成了具有中国特色的人工智能治理体系。

从治理思路上看，中国的人工智能治理更具柔性，更加灵活。希望以运转良好的系统性治理保证前沿科技探索的高速发展，同时防范发展所带来的隐疾。国家新一代人工智能治理专业委员会明确提出了中国人工智能

① 中央网信办：《习近平总书记关于网络强国的重要思想概论》，人民出版社2023年版，第40页。

② 《新一代人工智能发展规划》，中国政府网，2017年7月20日，https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm。

治理的八条原则，即和谐友好、公平公正、包容共享、尊重隐私、安全可控、共担责任、开放协作、敏捷治理，并将发展“负责任的人工智能”作为治理的目标。[①]

从治理过程来看，中国人工智能治理在执行过程中从不以限制技术创新与产业发展为目标，而是致力于打造人工智能参与各方都能够参与和发声的协作平台，实现发展带动治理，治理促进发展的良性互动。为了确保个人隐私、关键数据的安全问题，近年来中国先后颁布了《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》，以及各种行政法规和地方规章。但在处理生成式人工智能等新兴敏感技术创新时，又始终保持着开放性与灵活性，采用分级监管与分行业部门监管相结合的方式，在保证生成式人工智能技术安全使用的同时，尽量避免对该领域产业发展造成过大干扰。[②] 2023 年 10 月，在第三届“一带一路”国际合作高峰论坛上，中国领导人提出了《全球人工智能治理倡议》，[③] 完整、全面、深刻地阐述了人工智能治理的“中国方案”，在全球范围内产生了积极反响。

除中国之外，日本、俄罗斯在人工智能治理方面的选择也展现出较为明显的发展目标导向路径的特点。作为一度在人工智能技术领域占据领先地位的国家，日本非常重视本轮人工智能技术，希望能够借助人工智能技术进步带动国内经济发展，并且有效缓解人口老龄化、少子化等重大社会问题。为此，日本积极设计人工智能发展战略，并在处理人工智能治理问题时采取了较为务实的思路。日本政府拒绝将人工智能治理作为强制性法律推行，而是侧重于出台人工智能治理方面的软性规则与行为指南。日本政府认为，在社会面临数字化变革的时代，以法律法规为核心的传统治理模式难以跟上创新的步伐。因此，日本政府主张将治理模式从传统的基于规则的治理模式转变为基于目标的治理模式，以引导公司等产业实体的发展。[④]

① 《新一代人工智能治理原则——发展负责任的人工智能》，科技部网站，2019 年 6 月 17 日，https://www.most.gov.cn/kjbgz/201906/t20190617_147107.html。

② 《生成式人工智能服务管理暂行办法》，中国政府网，2023 年 5 月 23 日，https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm。

③ 《全球人工智能治理倡议》，中央网信办网站，2023 年 10 月 18 日，https://www.cac.gov.cn/2023-10/18/c_1699291032884978.htm。

④ ［日］“我国人工智能治理的方式 1.1”，日本经济产业省网站，令和 3 年 7 月 9 日。https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_1.pdf。

俄罗斯的选择思路也与此类似，由于受到国内经济体制与市场环境的制约，俄罗斯在人工智能领域有限的资源禀赋难以得到有效发挥。为此，俄罗斯采取了一种非常特殊的发展思路，将有限的资源集中在与国家安全密切相关的少数领域，建立某种闭环的技术研究与应用场景，希冀在军事安全等局部方向上取得实质性突破。[①] 俄罗斯以《2030年前俄罗斯联邦国家人工智能发展战略》作为该领域所有工作的指导，在该框架内，结合技术发展的特定情形定期推出相应的治理方案，其中最具代表性的是2020年签署的《2024年前俄罗斯人工智能和机器人技术领域监管发展构想》，[②] 这也是较为典型的发展目标导向的行动方案。

总之，以中国为主要代表的国家群体选择发展目标主导路径参与人工智能治理活动。该路径的实质是在处理人工智能问题时，始终以追赶技术与产业前沿作为首要目标，希望能够充分利用人工智能技术潜能推动国家的跨越式发展。治理作为发展战略的组成部分，起到的是为跨越式发展提供基石的作用。因此，选择该路径的国家的治理活动常常通过采取规范、原则等弱约束的方式，随着技术发展不断做出调整，以相对灵活的姿态为发展进程保驾护航，谨慎推动产业发展，避免对人工智能发展形成过多限制。

（四）多元化进程与竞争性图景

由于文章篇幅所限，我们只能选取世界主要大国在人工智能治理方面的立场、思路和举措进行简要描述，以此粗线条勾勒人工智能治理的全景图。但这并不会影响本文的结论，目前所有国家在人工智能治理领域的选择都没有脱离本文理论框架所描述的三条基本路径，即使将研究对象扩大到很多当下尚未形成完整人工智能治理方案的国家，基本结构也不会在短期内发生重大改变。可以说，这种沿三条基本路径平行推动的多元化进程构成了全球人工智能治理的基本形态，多种治理路径之间的持续竞争也充分体现了国际无政府状态下的主权国家行为逻辑与当前人工智能技术发展规律之间持续碰撞的状态。

① 封帅：《建构非对称竞争优势的尝试：俄罗斯人工智能治理体系的结构与逻辑解析》，载《俄罗斯学刊》2022年第5期，第72-73页。

② ［俄］“关于批准《2024年前俄罗斯人工智能和机器人技术领域监管发展构想》的政府令（N 2129-r）”，俄罗斯联邦政府网站，2020年8月19日，https://www.zakonrf.info/rasporiazhenie-pravitelstvo-rf-2129-r-19082020/。

一方面，治理的多元化进程反映出当前主权国家对于人工智能技术发展规律，以及人与人工智能关系的理解尚未形成共识。表现为各国政府对于人工智能机遇与风险的边界判断存在较大差异，这种差异化判断与各国在国际体系中的位置结合在一起，塑造了现有的多元化进程。各国内部对于人工智能治理各有设计，但受制于人工智能系统结构与规律的限制，最终的政策选择往往殊途同归，逐渐向上述三条基本路径靠拢。从长期来看，这种发展趋势将造就全球范围内不同治理体系长期并存的局面。

另一方面，由于多元进程的长期存在，全球人工智能治理领域必将呈现出越来越严重的治理竞争图景。在国际关系领域的基本规律影响下，各主权国家在面对人工智能治理议题时往往都首先从自身的利益需求出发，在处理治理议题时，治理的有效性与国际竞争中的主导权原则将必然出现矛盾。在这种情况下，国别利益压倒全球利益将是国际体系的常态。各国都将力推自身所设计和执行的治理体系，努力将自身的治理规则变为国际规则，从而为自身创造更加有利的发展条件。在竞争持续推进的情况下，治理本身也将逐渐从目的变为手段，有很大可能会逐渐偏离治理的原初目标。

经过上述理论与案例的分析，可以得出一个基本结论，即从国家维度展开全球人工智能治理图景，将会呈现出以三条核心路径为基本架构的多元化进程。虽然不同路径之间常常建立起互动关系，也有一定的对话与合作，但在无政府状态的制约下，不同路径之间，不同国家之间将展开越来越激烈的治理竞争，竞争性图景将成为国家维度下全球人工智能治理的主要特征。

五、余论：推进人工智能全球治理——一条漫长的道路

2016年至今，在不到10年的时间里，人工智能治理已经成为全球政治活动的重要内容。事实上，由于人工智能技术本身的巨大影响力，当前最重要的国际行为体——主权国家积极参与人工智能治理进程，并且逐渐取得主导权是当前国际结构所衍生的必然结果。由于治理体系建设是国家行为体干预人工智能系统运行的主要支点，因此得到了各国的充分重视，成为其战略设计的重要组成部分。从积极方面来看，主权国家是人工智能

治理进程得以推进的关键力量，在国家行为体的参与下，治理进程推进速度不断加快。但其消极的一面也不可忽视，身处国际无政府状态中的主权国家，其所开展的治理活动会受到技术规律与国际关系规律的共同影响。作为这种影响的现实结果，从国家维度展开的人工智能治理进程呈现出非常清晰的竞争性图景，这场争夺治理主导权和国际规则的竞争恐将愈演愈烈。

然而，如果从梳理人类与人工智能关系的理想视角来看，智能时代人类所需要的并不是国别维度的治理竞争，而是基于共同认知与共同目标基础上的人工智能全球治理。国家维度的人工智能治理进程全球图景研究展示给我们的是一幅各区域存在明显治理落差的图景，更是一幅很多国家试图以治理为武器实现获取权力目标的竞争图景。在这种状态下，一旦人与人工智能的关系在短时间内出现重大变化，现有的治理格局难以进行有效应对和处理。即使没有技术层面的剧变，治理竞争一旦进一步恶化，也将产生严重的负面影响。我们深知，在国际无政府状态下，现有状况的形成具有一定的必然性，该结构也具有强大的韧性，想要改变并不容易，推动人工智能全球治理必然是一条漫长且艰难的道路。我们在此分析和展示国家维度的人工智能治理进程，最根本的目的是要让所有关心人工智能治理的人们清晰地了解到我们所面临的挑战，集结知识与智慧，找到改变现状的合理路径，推动真正公平、敏捷、高效的人工智能全球治理体系的建构，这也应该成为所有人工智能治理议题研究者的共同目标。

Two-way Empowerment of Developing New Quality Productive Forces and Building the "Belt and Road": Logic, Theoretical Framework and Practice

Han Yonghui; Li Siyi; Cheng Hao

Abstract: At present, the world is navigating a new period of turbulence and transformation, which has reignited the trend of unilateralism, thereby amplifying the external risks for developing new quality productive forces and pursuing high-quality development under the Belt and Road Initiative. This paper first delineates the intrinsic connections between the development of new quality productive forces and the collaborative construction of the "Belt and Road". Subsequently, it examines the mechanisms that mutually enhance development between the two, constructs a theoretical framework for two-way empowerment, and, based on these findings, explores practical pathways for the development of new quality productive forces and the two-way empowerment within the Belt and Road Initiative.

Key words: New Quality Productive Force Strategy; Belt and Road Initiative; High-quality Development

Authors: Han Yonghui, Professor, Guangdong Institute for International Strategy, Guangdong University of Foreign Studies; Guangdong Research Center for Xi Jinping Thought on Socialism with Chinese Characteristics for a New Era at Guangdong University of Foreign Studies.

Li Siyi, Research Assistant, School of Advanced Translation, Guangdong University of Foreign Studies;

Cheng Hao, Master Candidate, Guangdong Institute for International Strategy, Guangdong University of Foreign Studies.

Strategy and Power: The Role and Impact of Commercial Entities in the Militarization Process of Artificial Intelligence

Zhang Luyao; Lu Chuanying

Abstract: The involvement of commercial entities in the militarization of artificial intelligence (AI) is an important and special phenomenon. These new actors do not only fully participate in the entire process of research development, applica-

tion, and deployment in military AI, but also play a significant role in enhancing defense capabilities and strategic deployment of countries, affecting the overall direction of the militarization process of AI in a state. Case studies from the United States, Russia, Israel, and France show that the different roles of commercial entities in the military structure and process have shaped each state's AI militarization pace differently, and have impacted the innovation in national defense military capabilities and deployment strategies. States need to adjust their relationships with commercial entities according to their actual circumstances, complementing the shortcomings of the traditional defense industry system and stimulating the vitality of the military system, while guiding a harmonious interaction between commercial strategies and national security strategies, so as to jointly contributer to a responsible applications of military AI.

Key words: Artificial Intelligence Militarization; Big Techs; Military−Industrial Complex; Defense Innovation

Authors: Zhang Luyao, Ph.D Candidate, School of International Relations and Public Affairs, Fudan University; Lu Chuanying, Senior Fellow and Deputy Director, Institute of Public Policy and Innovation, Shanghai Institutes for International Studies

Comparison of Artificial Intelligence Governance Model Between the EU and the US

Yan Shaohua; Yang Zhao

Abstract: With the extension of global competition, artificial intelligence has turn into a pivotal technology related to national security while the governance of artificial intelligence has become a priority in both domestic and global arenas. The European Union (EU) and the United States (US) are two leading technical actors in artificial intelligence, which are playing critical roles and leading two models in governance. The EU's governance model features "hard regulation" based on rights and risks, emphasizing the establishment of a comprehensive framework of governance through laws and regulations, in order to enhance its "normative power". In contrast, the US' governance model is characterized

by "soft regulation" relied on "soft tools", for instance, the presidential order, encouragement of "industry self-governance" and "self-regulation", in order to maintain its technical leadership. The divergence of two governance models is manifested in three aspects, including the usage of governance tools, the design of governance institutions, and the participation of private actors, which can be explained by three major factors, namely policy traditions, influence of interest groups, and ecology of the artificial intelligence industry. In fact, these two models do not have substantial conflicts, on the contrary, both are appropriate approaches in line with their unique circumstances. As artificial intelligence develops, it is also urgent for China to develop its own governance model through seeking a balance between the EU model and the US model.
Key words: European Union; United States; China; Artificial Intelligence Governance
Authors: Yan Shaohua, Associated Professor, Institute of International Studies, Fudan University; Yang Zhao, Ph.D Candidate, School of International Relations and Public Affairs, Fudan University

An Analysis of the Construction of Artificial Intelligence Global Governance Regime Complex

Gui Changni

Abstract: Artificial intelligence (AI) has risks in internal security, application security, national security and geopolitical security, which inform different governance paths. The reflected characteristics of the regime complexity lead to the aggravation of the dilemma in artificial intelligence global governance. The root of the dilemma is that the governance mechanism fails to reflect the characteristics of the various artificial intelligence security risks. Moreover, the governance pattern fails to reflect the power changes and interactive relationships of the actors. The leading governance concept also fails to reflect the development trend of AI technology. The regime complex theory is introduced to deal with the dilemma of artificial intelligence global governance, and to construct an artificial intelligence global governance regime complex so as to implement modular dynamic

governance based on stratification; carry out collaborative governance relying on multiple entities based on the center; and build a coordinated responsibe governance among major powers based on mutual trust.

Key words: Artificial Intelligence; Global Governance; Governance Dilemma; Regime Complex

Author: Gui Changni, Senior Engineer, China Information Technology Security Evaluation Center

Global AI Governance: A Diversified Process and Competitive Landscape

Feng Shuai; Xue Shikun; Ma Yiruo

Abstract: The rapid development of artificial intelligence (AI) technology has had a significant impact on the international system. The national perspective can serve as an important lens for analyzing the current global AI governance process. From the national perspective, the global AI governance landscape unfolds as a diversified process structured around three core pathways: the power−oriented path, represented by the United States as the sole example, is characterized by the maintenance of hegemony in the field of intelligence; the security−oriented path, represented by the European Union and other countries, prioritizing regulation; the development−oriented path, represented by countries such as China and Japan, combines governance tools with strategic development in the hope of achieving technological and industrial catch−up. Under anarchy, diversification and competition will emerge as the main features of global AI governance from the national perspective.

Key words: AI Governance; Global Landscape; National Dimension; Diversified Process; Competitive Landscape;

Authors: Feng Shuai, Associate Professor, Institute for International Strategic and Security Studies, Shanghai Institutes for International Studies; Xue Shikun, Master Candidate, Shanghai Institutes for International Studies; Ma Yiruo, Master Candidate, Shanghai Institutes for International Studies