



AI-AIG
 清华大学人工智能国际治理研究院
 INSTITUTE FOR AI INTERNATIONAL GOVERNANCE
 TSINGHUA UNIVERSITY



全球人工智能治理与中国方案

(征求意见稿)



AI-AIG
 清华大学人工智能国际治理研究院
 INSTITUTE FOR AI INTERNATIONAL GOVERNANCE
 TSINGHUA UNIVERSITY



目录

一、当前全球人工智能治理的现状与挑战 1

- (一) 人工智能发展的基本情况 1
- (二) 人工智能全球治理的进展 3
- (三) 人工智能国际治理面临的挑战 6

二、中国实践为全球人工智能治理注入动力 7

- (一) 发展中国特色人工智能治理体系，引领国际治理体系变革 8
- (二) 加快人工智能科技创新和产业应用，推动全球人工智能发展 9
- (三) 发出全球人工智能治理倡议，提出全球治理“中国方案” 11

三、积极构建全球人工智能共同治理体系的建议 13

- (一) 求同存异，持续沉淀基于共识的全球人工智能治理理念 13
- (二) 承继创新，着力构建开放包容的全球人工智能治理体系 14
- (三) 循序渐进，分步打造综合系统的全球人工智能治理框架 17

参考文献 19

全球人工智能治理与中国方案

（征求意见稿）

清华大学人工智能国际治理研究院、阿里巴巴集团

历经六十多年发展，人工智能已从量变接近质变临界点。近年来，随着深度学习技术的持续突破，人工智能领域正在历经一场由生成式人工智能引领的爆发式发展。其中，又以生成式大语言模型（LLMs）技术的突破性发展为节点，人工智能科技出现从专用化走向通用化的可能性，标志着弱人工智能向强人工智能（通用人工智能）跃迁。

人工智能技术展露深入千行百业的禀赋，被视作推动现代产业高质量发展的重要支撑和孵化新质生产力的重要引擎，或将引发各领域体系性变革。例如，在经济领域，变革传统产业，创造未来产业。在社会领域，变革生活、就业、学习等方式。在文化领域，改变内容生成和传播范式。在科技领域，催生科研新范式，开辟全新科学时代。其还可能超常规改变国家经济社会安全水平，改变国家间力量对比，影响全球地缘政治和竞争格局。同时，人类还共同面对人工智能发展带来的伦理、秩序、安全挑战。

当前，全球人工智能治理方兴未艾，不同的国家、国际组织、智库、专业组织、企业等均提出有关人工智能治理的倡议、立法、报告、主张。在中国，政府、产业、学界和社会各界经历一年多的探索，对以生成式大语言模型为代表的生成式人工智能的发展和治理，形成了基础共识和治理框架，并将随大模型技术的持续演进（如多模态能力）和应用场景的不断丰富而持续迭代，叠加生动活泼的产业实践，以为国际人工智能治理贡献智慧、注入动力。

一、当前全球人工智能治理的现状与挑战

（一）人工智能发展的基本情况

当前正处于人工智能发展的关键期，国际竞争激烈。美国处于技术与产业应用的双领先地位，中国、欧洲虽发展迅速，但基础模型能力仍处于技术追赶阶段，在打造应用生态方面也并无显著优势。中、美、欧“你追我赶”的竞争态势，一定程度上塑造了当

下国际治理的基本格局。但值得注意的是，人工智能国际治理的基本面并不完全取决于各主要经济体的战略、政策选择或相互竞争关系，而是更为本质地取决于现实的技术发展水平，尤其是产业实际应用深度。

在普通用户端，大模型应用尚未走出通用型工作场景，而嵌入大众的日常生活。人工智能在技术范式上实现了从“小模型+判别式”向“大模型+生成式”的转变，从传统的人脸识别、目标检测、文本分类，升级至如今的文本/图像/语音/视频/3D 数字人生成。然而，当下的大模型最为成熟的生成式应用依然集中在“语言生成”方面，产品形式表现为聊天工具或者辅助搜索工具，主要用户群体为希望进行办公效率提升的脑力劳动集中型人群。也即大语言模型的成熟产品形态依然是借助大模型产品提升办公效率的商业用户端场景，这与传统移动互联网时代大量普通用户端应用的普及形成鲜明反差。

在商业用户端，大模型应用正处在从通用工作场景走向核心工作场景的初步阶段。大语言模型技术“偏科”于语言生成，这导致其早期的商业用户端应用主要表现为对“通用环节”工作流的优化和效率提升，特别是在企业内部搜索、文本书写、代码生成、法律助手环节上表现优异，而对于行业内既有的“核心 workflow”——如偏情感互动以及个性化创新的大客户销售、医疗诊断、投资分析、学术创新，或重追求稳定性、低容错率的工业制造的核心环节——大语言模型嵌入度有限。当下，基础模型能力提升、算力成本下降正在带来“性价比”飞轮效应，行业场景中沉睡的高质量数据的结构化清洗、引入已经逐步成为一项重要产业，可以预见大模型在商业用户端的应用场景将从办公、编程、营销等通用环节走向更为核心的生产环节。

表 1：人工智能未来 3-5 年发展中的几个关键问题^①

关键问题	详细阐述
原生多模态生成式大模型	文字数据只占全世界数据量的 10% 不到，更多的数据以图像、视频方式存在。 目前的文生图、文生视频，并非真正意义上的多模态。 终极形态是类人脑多模态大模型。
Agent 能力	场景 1：AI 手机，AI 端工具调用所有 APP，改变人机交互模式，交互界面从图形（GUI）为主走向语言为主（LUI）。 场景 2：用户的“个人助手”“高级助理”。

① 综合北京智源人工智能研究院、清华大学等多场公开/闭门学术会议专家发言综合整理。

软硬件组合优化及具身智能	人工智能的进步不仅体现在算法和模型的优化上，更在于计算能力的提升。具身智能被视为人工智能的终极形态。
AI For Science	用大模型来进行科学技术的突破，如分子结构预测模型。
安全可靠	定义短时间内难以取得一致。分类维度包括：人类生存性风险或一般性风险；人工智能系统本身的风险，或系统赋能型风险。
产业界持续主导	斯坦福大学《2024年人工智能指数报告》显示，2023年，产业界产生了51个著名的机器学习模型，学术界只贡献了15个，产业界继续主导人工智能前沿研究。

总体来看，以生成式人工智能为代表的本轮人工智能浪潮，总体处于“工具化末期、产品化初期”阶段。回望过去，数字技术落地为工具、进化为产品，到完成市场定位、获得消费者认可，并最终转动商业飞轮，无法一蹴而就。生成式人工智能技术架构从理论创新提出到工程化落地，距今不过约5年，以 ChatGPT4.0 和 Sora 发布为节点，人类两次感受到“大模型+生成式”的冲击，至今不过1年。当前，大模型行业距离完美的产品-技术契合（Product Technology Fit, PTF）、产品-市场契合（Product Market Fit, PFM），尤其是解决性能-成本匹配问题，催生“AI原住民”，还有相当的距离。

（二）人工智能全球治理的进展

第一，“发展治理”与“安全治理”需求并发。2023年3月，OpenAI公司旗下智能对话聊天机器人 ChatGPT4.0 新版本发布，其出色的自然语言生成能力引起了广泛关注，国内外掀起了一场大模型创新创业浪潮。在海外，GPT、Gemini、Llama、Claude、Mistral 等多款大模型如雨后春笋般涌现，版本不断迭代，形成了以 Open AI、谷歌、Meta、Anthropic、Mistral AI 为代表的基础模型研发第一梯队。在国内，阿里、百度等传统数字科技平台企业推出的通义、文心一言等大语言模型，初创企业智谱、月之暗面等推出的智谱清言、Kimi 也取得了较好的市场效果，形成了国内自研大模型第一梯队。行业千帆竞逐，“发展型治理”成普遍关切。受算力、数据、算法人才、风险投资等生产要素不同程度上的短缺所制约，相较于美国，欧洲以及包括中国在内的广大发展中国家在 AI 核心能力方面，正处于并可能在未来一段时间内继续处于“跟跑”状态。中美比较而言，中国在 AI 高端人才数量、AI 基础算法创新、AI 底座大模型能力（大语言模型、文生图模型、文生视频模型）、底座大模型训练数据、底座大模型训练算力等方面，都与美国存

在一定的差距^①。为解决这一问题，中国、欧洲、中东、东南亚等国家和地区相继发布国家基础设施及 AI 投资计划，聚焦可持续算力、高质量数据集等大模型研发的核心生产要素补齐，同时围绕人才、资本、监管等制度要素进行合理配置，逐步形成了一套“发展型治理”体系。与此同时，“安全治理”备受关注。尽管目前大模型技术的应用尚且处于初步阶段，但受“虚假新闻”等现实暴露风险影响以及远景不确定的预期塑造，政策和公众侧对于拟人化、生成式的大模型可能带来的“虚假信息”“深度伪造”“隐私侵权”“版权侵害”甚至潜在的“人机关系失调”、“就业替代”、“分配不公”感到担忧^②，基于风险防控的风险焦虑同时出现。欧盟《人工智能法》落地并逐步付诸实践。英国、韩国先后举办以 AI 安全（AI Safety）为主题的“布莱切利安全峰会”和“2024 人工智能首尔峰会”。美国推出了 AI 安全的行政令。英国、美国、日本、韩国和加拿大等国纷纷成立人工智能安全研究所。中国颁布《生成式人工智能服务管理暂行办法（试行）》，学界也出现了为人工智能立法的思考^③。

第二，“为发展而治理”逐渐上升成为核心共识。人工智能发展至今，世界主要经济体尤其是大国都认识到人工智能对于经济增长、国家安全和福祉至关重要，同时认识到不受约束的人工智能可能带来的风险，无一例外地主张发展与安全并重，一方面积极促进人工智能技术开发与产业应用，另一方面积极探索人工智能治理。尽管受技术实力、产业禀赋差异影响，以及政治、历史、社会、文化、价值观等一系列要素制约，主要经济体在对待发展与安全的优先级方面存在差异，但为发展而治理、基于发展开展治理正在成为国际社会的普遍的共识。以欧盟为例，欧盟长期对数字经济采取全面而严格的监管，一定程度上引起“反噬”讨论。在此轮人工智能浪潮中，欧盟一方面以《人工智能法》建立一套基于风险分类的人工智能规制体系，对高风险 AI 系统实行严格的预审核，治理路径上强调政府直接监管与法律硬约束，另一方面在法、德等国的坚持下在立法最后关头微调立场，对《人工智能法》做宽松回撤，将大模型风险独立出来对待，“既要降低风险，又要构建一个促进人工智能领域投资和创新的框架”。相较于欧盟整体，法、德等国展露出更为积极的拥抱 AI 的态度。据德国数字化和交通部部长维辛（DR.

① 《人工智能与智能计算的发展》，孙凝晖，十四届全国人大常委会专题讲座第十讲讲稿。

② Pew 民调显示：对人工智能担忧多于兴奋的美国人比例，从 2022 年的 37% 增长到了 2024 年的 52%。Ipsos 民调显示：2023 年，认为人工智能在未来 3-5 年内对他们的生活产生巨大影响的比例，从 60% 上升到 66%。

③ 出现了两部“人工智能法”学者建议稿。

VOLKER WISSING) 介绍, 2023 年, 德国诞生了 500 家 AI 初创企业, 60 万家德国企业已经在其工作中嵌入了 AI, 人工智能已经成为德国数字政策的焦点^①。法国驻华大使白玉堂 (H.E. Bertrand Lortholary) 介绍, “法国正不断加大对人工智能人才、基础设施、资金、利用与治理等方面的投入”^②。布鲁塞尔效应回撤的效果明显: 统计显示, 2023 年, 欧盟以 23 个优质开源模型领先中国 (15 个), 而仅落后于美国 (61 个)。法国 Mistral AI 在欧一枝独秀^③。

第三, 全球安全治理尚且处于早期阶段。其一, 安全风险认知尚处初步阶段。尽管全球热议人工智能安全治理, 各国国内有关安全风险防控的立法提案不断, 但当前人类对于前沿人工智能 (Advanced AI) 的工作机理、风险水位、潜在风险的认知尚处在初步, “人工智能安全” 的内涵与外延、治理的手段与工具, 都将随着技术的演进以及人类对于人工智能安全科学研究的不断加深而持续迭代。其二, 尝试建立敏捷治理的风险防控体系。考虑到大模型技术的快速演进和广泛的场景适用, 主要经济体均采取动态、弹性的框架来开展治理, 着眼模型开放透明度以及以负责任的方式提供大模型产品服务两大方面^④。今天, 各国普遍主张平衡风险与机遇、多利益相关方^⑤参与、分布式和迭代治理、自愿性框架与立法相结合、透明度和问责制等。值得注意的是, 由于多国纷纷提出自身的治理框架, 框架之间的互操作性得到重视, 联合国人工智能高级别咨询机构于 2023 年 12 月发布《为人类治理人工智能》临时报告明确的下步重点工作之一, 即为 “通过在世界范围内 (联合国) 认可的全球人工智能治理框架, 加强世界各地新出现的治理工作的互操作性及其在国际规范中的基础”^⑥。

① Exclusive interview with DR. VOLKER WISSING. Visionsforeurope, <https://visionsforeurope.eu/v4e-articles/exclusive-interview-with-dr-volker-wissing/>, 2024 年 6 月 1 日最后访问。

② 《CCG 与德法使馆联合举办研讨会, 热议中欧视角下的人工智能治理》, <http://www.ccg.org.cn/archives/85101>, 2024 年 6 月 10 日最后访问

③ THE AI INDEX REPORT Measuring trends in AI, <https://aiindex.stanford.edu/report/>

④ Global AI Law and Policy Tracker, IAPP, https://iapp.org/media/pdf/resource_center/global_ai_law_policy_tracker.pdf, 2024 年 5 月最后访问

⑤ 以美国 NIST 风险框架 1.0 版本为例, 其所列举的利益相关方包含 14 种, 涵盖了整个人工智能的大生态, 包括设计方 (Design)、开发方 (Development)、部署方 (Deployment)、运营和监控方 (Operation and Monitoring), 测试、评估、验证和确认方 (Test, Evaluation, Verification, and Validation, 简称 TEVV), Human Factors、领域专家 (Domain Expert)、人工智能影响评估方 (AI Impact Assessment)、采购方 (Procurement)、治理和监督方 (Governance and Oversight), 此外, 还可能包括第三方实体 (Third-party entities)、最终用户 (End users)、受影响的个人与社区 (Affected individuals/communities)、其他 AI 行为体 (Other AI actors) 以及大众 (The general public)。

⑥ Interim Report: Governing AI for Humanity, AI Advisory Body of United Nations, 2023 年 12 月, https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf, 2024 年 6 月 7 日最后访问。

（三）人工智能国际治理面临的挑战

第一，前沿人工智能的技术挑战。前沿人工智能国际治理体系已经有了初步的形成基础，即国际社会均认同前沿人工智能对于全人类的巨大潜在风险，如何防范技术失控和滥用成为最核心的国际议题。

其一，失控风险主要指的是在开发、设计、部署和应用过程中，由于技术缺陷导致人工智能失控所造成的风险。特别是在前沿人工智能领域，在人类无法完全理解并时刻掌控人工智能大模型等技术之前，如果将其应用于各种具体场景，特别是与物理世界连接，由此产生的风险可能是难以预期、无法溯源的，并且可能带来系统性、灾难性的后果，如当前快速发展的自动驾驶、机器人、具身智能等，在与大模型能力结合带来性能提升的同时也面临失控风险。其二，技术滥用风险指的是除一些人为操作不当和管理疏忽所导致的滥用外，主要在于防止军事领域的人工智能滥用，以及防止恐怖分子、极端组织对于人工智能技术的滥用。这些都需要国际各方尤其各国政府之间达成共识与合作。

人工智能正成为当今世界最具颠覆性的技术力量，放任相关技术无限制发展的重大风险已引起国际社会普遍担忧，正是因为没有一个国家能够独善其身、免于这些风险，成为推动国际共识的最大动力。

第二，前沿人工智能的国际治理挑战。尽管前沿人工智能已经有了形成国际治理体系的动力，但同时也面临着极大的挑战。

其一，治理机制变革与技术发展缺乏同步性。目前，已有国际组织和治理机制很难跟上新兴技术的发展速度，人工智能国际治理呈现分散、低效特征。主要原因有两个方面：（1）由于技术水平、产业结构、人才储备、管理制度等方面的不同，不同国家在前沿人工智能应用和管理重点均存在差异，甚至存在部分国家的措施和重点相互冲突的现象。（2）人工智能作为革命性新兴技术，可能会使一个国家在技术革命中成为绝对赢家。尤其人工智能大模型等前沿人工智能在实际应用中展现出惊人潜力后，许多国家在防范技术风险的同时，将更多关注点放在制造能被国际认可、但又有利于本国人工智能产业发展的政策洼地，由此导致国际分歧不仅没有弥合，反而有愈发加深的风险。

其二，发达国家和发展中国家技术鸿沟难以解决。国际货币基金组织（IMF）等机

构的研究显示，未来发展中国家与发达国家的差距会越来越大^①。AI所引发的发达国家和发展中国家技术鸿沟主要表现在：AI技术的投资将会更加青睐已经“实现自动化”的发达经济体；AI技术不断替代发展中国家在低生产成本等方面的传统优势；AI技术对发达国家和发展中国家的赋能存在极大差别；发展中国家由于产业发展薄弱，在全球人工智能治理话语体系的影响力有限等等。

其三，实际治理系统复杂，共识处于初步阶段。人工智能技术的复杂性和通用性决定了其治理的系统性与挑战性。以生产要素为视角，涉及到算力、数据、人才、技术、市场、资本等多种要素的组合。以模型产业链为视角，涉及算力、模型、应用等相关市场，涉及开发、部署、使用等多个流程环节。以社会影响为视角，涉及就业替代、人机关系、能源可持续、大国竞争等议题。当下，若仅从语词层面分析，前沿人工智能治理并不存在显著分野，相关表述相通甚至可互换，而这恰恰是因为技术发展、产业发展尚在初步，核心风险问题处于“待定义”状态，全球也因此尚未能形成一套被普遍接受的风险治理体系和框架，距离可操作的治理细则落地还有很长一段路要走。目前，从治理进度看：整体尚处于共识磨合期、机制构筑期、行动计划期，实际治理动作并未“大步向前”^②。横向治理看：为对过往弱人工智能系统治理、网络空间治理议题的总集和“刷新”，集中表现为对数据、算法等要素的规制（见表格1）。纵向治理看：集中表现为对医疗、交通、金融等重点风险域的用例规制。治理机制看：以既有机构、机制为主，新增新设机构、机制为辅。联合国作用重要，多边力量积极参与。此外，虽然统筹发展与安全是共识，但各主要经济体对于发展、安全有着并不完全一致的优先偏好，对于发展与安全的内涵和外延，也有着不尽相同的定义，人工智能治理面临创新与监管的平衡困境。凡此种种，导致当前全球人工智能治理合作既无比重要、又困难重重。

二、中国实践为全球人工智能治理注入动力

作为全球人工智能赛道的领跑者之一，中国一直以来高度重视人工智能发展和风险防范，不仅在人工智能技术创新和产业应用方面贡献了重要力量，还通过一系列国内治

^① Cristian Alonso Siddharth Kothari and Sidra Rehman. How Artificial Intelligence Could Widen the Gap Between Rich and Poor Nations, IMF, <https://www.imf.org/zh/Blogs/Articles/2020/12/02/blog-how-artificial-intelligence-could-widen-the-gap-between-rich-and-poor-nations>

^② 以联合国层面治理看，联合国秘书长设立的人工智能咨询机构，于2023年12月发布《为人类治理人工智能》临时报告，今年9月将发布正式报告。目前尚未正式开展相关实际治理动作。

理创新和国际倡议在人工智能治理领域起到了重要的示范和引领作用，展现了科技大国的责任担当，为人工智能全球治理注入了重要动力。

（一）发展中国特色人工智能治理体系，引领国际治理体系变革

中国国家主席习近平多次强调要重视人工智能发展，营造创新生态，重视风险防范，为中国兼顾人工智能创新发展和风险防范的治理模式指明了方向。近年来，中国人工智能治理机制不断完善，具有多元主体、多维共治、多种工具、敏捷协同等特征的中国特色人工智能治理体系初见雏形。

第一，规划先行，完善人工智能治理的顶层设计。2017年，中国发布《新一代人工智能发展规划》，赋予新一代人工智能“国际竞争新焦点、经济发展的新引擎、社会建设的新机遇”的定位，提出了“加强前瞻预防与约束引导，最大限度降低风险”的战略要求和“科技引领、系统布局、市场主导、开源开放”的基本原则。2019年，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》，围绕发展负责任的人工智能主题，强调了和谐友好、公平公正、包容共享、尊重隐私、安全可控、共担责任、开放协作、敏捷治理八项具体治理原则。2020年《国家新一代人工智能标准体系建设指南》明确了安全、伦理双轮驱动的标准体系建设方向，安全部分包括基础安全，数据、算法和模型安全，技术和系统安全，安全管理和服务，安全测试评估，产品和应用安全等六大部分，伦理部分聚焦医疗、交通、应急救援等特殊行业。通过发展规划、治理原则、建设指南形成了较为完整的人工智能治理体系顶层设计。

第二，平衡发展安全，加快中国特色治理实践。在促进发展方面，2019年，中国共产党中央全面深化改革委员会第七次会议提出“促进人工智能与实体经济深度融合”。2019年至2021年，中国科技部在全国布局了18个“国家新一代人工智能创新发展试验区”，支持开展技术示范、政策试点和智能化基础设施建设，推动打造创新策源地和产业高地。2022年，科技部等六部门印发《加快人工智能场景创新和示范应用》，建设近千个应用场景，探索场景驱动的人工智能赋能经济社会发展新模式。此外，《深圳经济特区人工智能产业促进条例》《上海市促进人工智能产业发展条例》等地方条例陆续推出，旨在推进人工智能与经济、生活、城市治理等领域深度融合，促进鼓励人工智能创新发展。在风险防范方面，2021年《新一代人工智能伦理规范》涉及管理、研发、供应、使用四大环节。在算法设计、实现、应用方面，要求提升透明性、可解释性、可理解性、可靠性、可控性，增强人工智能系统的韧性、自适应性和抗干扰能力；在数据采集和算

法开发方面，要求提升数据质量，考虑差异化诉求，避免可能存在的数据与算法偏见；在人工智能产品和服务提供方面，要求充分尊重和帮助弱势群体、特殊群体，并根据需要提供相应替代方案。此外，还要求保障人类拥有充分自主决策权，确保人工智能始终处于人类控制之下。2022年，《关于加强科技伦理治理的意见》提出“增进人类福祉、尊重生命权利、坚持公平公正、合理控制风险、保持公开透明”等治理原则。

第三，依托原有治理体系，加强横向、纵向规制建设。横向规制方面，一是以数据和算法安全为核心的双轨规制制度完备。基础法律方面，《电信条例》《电信法》《网络安全法》《数据安全法》《个人信息保护法》《科技伦理审查办法（试行）》《生成式人工智能服务管理暂行办法》《数据出境安全评估办法》《关键基础设施安全保护条例》，擘画了AI时代的治理之基，尤其为规范人工智能应用中的数据安全和隐私保障提供了合规指引。二是部门规章方面，针对人工智能的特殊性，《算法推荐管理规定》《深度合成管理规定》提出了算法备案、数据安全审核以及训练数据治理等制度要求，进一步完善了中国人工智能治理的监管体系。三是标准建设和行业合规实践方面，《生成式人工智能服务安全基本要求》《生成式人工智能服务内容标识方法》相关标准为行业合规提供了方向，行业以模型备案为主要抓手，基本落实了语料安全、模型安全、服务安全等大体框架，涵盖了中国伦理和立法相关的要求，且有相关的具体比例^①。纵向规制方面，在生物特征识别、自动驾驶、智慧医疗、金融等重点AI+领域，治理水位达到世界领先水平。以生物特征识别为例，中国已围绕人脸识别先后发布了数据安全标准、系统技术要求等，并围绕人脸、步态、基因、声纹等发布了安全要求。以自动驾驶为例，有《智能网联汽车道路测试管理规范（试行）》《智能网联汽车道路测试与示范应用管理规范》等规定。智慧医疗方面，有《人工智能医疗器械注册审查指导原则》《人工智能辅助诊断技术管理规范》《人工智能辅助治疗技术管理规范》《人工智能辅助诊断技术临床应用质量控制指标》等。

（二）加快人工智能科技创新和产业应用，推动全球人工智能发展

第一，领先的中国互联网行业为人工智能提供了丰富的数据、算力资源和多样的应用场景。无产业支撑，治理将成无源之水。公平公正、生动活泼的治理格局，离不开百花齐放的产业生态。中国全功能接入互联网30年，基于中国超大规模的统一大市场优

^① 如要求测试题集不低于4000题，合格率不低于95%。

势和超前的数字基建，叠加制造业结构、高素质劳动者队伍，孵化出阿里、腾讯等一流互联网企业，塑造了全球互联网行业双巨头格局，并深刻影响全球数字治理。中国互联网大潮中，行业不仅积累了数据的量，更积累了以“芯片+云服务”为核心的数据计算力，以“数据+算法”为核心的数据产品力，以“平台+用户”为核心的数据服务力。数字人才丰富、数字资本充裕、数据能力突出，科技平台作为创新要素的连接器、新技术新应用的孵化器、“人工智能+”的赋能器，为中国人工智能快速发展，奠定了坚实的算力、数据、算法基础，为行业从模型走向应用提供了条件。

第二，人工智能核心产业发展迅速，顶尖模型不断取得突破。斯坦福人工智能研究所（HAI）发布的《2024年人工智能指数报告》^①显示，截至目前，中国人工智能综合指标位列全球第二，顶级智能模型数量在全球位列第三（15款），仅次于美国（61款）和欧盟（21款），专利数量第一。截至2023年底，中国人工智能核心产业规模接近5800亿元，核心企业数量超过4400家，居全球第二^②；到2024年4月，中国已累计完成117项模型备案，^③迎来了行业模型的备案热潮。

专栏：以阿里巴巴为例，看中国AI产业的国际竞争力

当前，懂AI、用AI成为新时代定义科技公司的新token，AIGC的氛围感、生态感正在孕育，阿里巴巴不仅喊出了“All In AI”的口号，还落在了具体行动上。例如，基于互联网时代的长期能力经验积累，进入生成式人工智能时代以来，阿里巴巴既是大模型技术与产品的研发方，也是重要的对外服务提供方、部署方，还是大模型技术与产品的重要应用方之一。

阿里在算力服务供给、模型研发和生态应用三大方面，全面拥抱大模型时代，旗下“通义”系列模型产品，今年5月发布的通义2.5版本（闭源版本），中文性能超越GPT4Turbo，系国内首次。今年6月发布的Qwen2（开源版本），包括了5个尺寸的预训练和指令微调模型，即Qwen2-0.5B、Qwen2-1.5B、Qwen2-7B、Qwen2-57B-A14B以及Qwen2-72B，综合表现超过Meta开源版本Llama3和国内目前所有的闭源模型。

① <Artificial Intelligence Index Report 2024>, <https://aiindex.stanford.edu/report/>

② 赛迪智库人工智能产业形势分析课题组：《2024年我国人工智能产业展望》，2024年1月。

③ 《国家互联网信息办公室关于发布生成式人工智能服务已备案信息的公告》，https://www.cac.gov.cn/2024-04/02/c_1713729983803145.htm

第三，人工智能加快赋能千行百业，创造大量智能向善的“中国案例”。中国拥有全球通讯基站覆盖率最高的通信网络、全球最为广泛使用的电子支付系统和全球最完整的工业体系，不仅是人口大国、数据大国，更是市场大国、应用场景大国，这些优势为中国人工智能的落地应用提供了广阔前景。2024年，中国进一步开展旨在大力培育人工智能新产品和新服务的“人工智能+”行动，促进人工智能与实体经济深度融合，推动人工智能赋能新型工业化。目前，“人工智能+”行动已在工业、医疗、金融、教育、电商、交通等诸多领域落地实践，不断促进人工智能技术更好造福人类。根据工业和信息化部赛迪研究院统计，到2035年，生成式人工智能产业对中国经济的贡献有望突破30万亿元人民币。^①

专栏：多个中国案例入选首批联合国智能向善案例集

2024年5月30日，人工智能向善全球峰会（AI for Good Global Summit）于瑞士日内瓦国际会议中心举办。作为联合国在AI领域的旗舰平台，峰会旨在推动AI围绕健康、气候、包容性发展、清洁能源等的应用。此次峰会公布全球首批40个AI for Good案例集，其中多项中国案例入选，并获评优秀案例。

例如，阿里巴巴达摩院东南亚语言大模型SeaLLMs针对东南亚地区语言繁多且语料不足、缺乏实质性语言技术支持的现状，采用多种定向增强技术，包括扩展专用词表处理非拉丁字符语言，更好地捕捉地区间语言的细微差别，在泰语、越南语、印尼语等广泛的语言任务上表现优越。与众多现有的解决方案相比，SeaLLMs致力于降低非英语社区的人群获取尖端AI技术的门槛，尊重和反映当地的文化、习俗、风格偏好和法律规范，助力防范文化同质化、保留语言多样性。

再如，商汤科技联合上海银行创新打造商汤如影AI数字员工“海小智”和“海小慧”助力银发群体跨越“数字鸿沟”，为不习惯页面操作相对复杂APP的老年用户群体降低了使用门槛，超写实高精形象也让老人能感受到有温度的陪伴式交流。

（三）发出全球人工智能治理倡议，提出全球治理“中国方案”

2021至2023年，中国连续提出全球发展倡议、全球安全倡议和全球文明倡议，为解决全球发展问题、安全问题和文明交流互鉴问题提供了切实可行的路径和方案。以此

^① 赛迪智库人工智能产业形势分析课题组：《2024年我国人工智能产业展望》，2024年1月。

为基础，中国作为人工智能发展与利用的重要力量，在参与全球人工智能治理体系构建、推动人工智能国际治理合作上，始终主张各国应秉持共同、综合、合作、可持续的安全观，坚持发展和安全并重的原则，通过对话与合作凝聚共识，构建开放、公正、有效的治理机制，促进人工智能技术造福于人类，推动构建人类命运共同体。围绕这一主张，中国积极支持在联合国框架下协调国际人工智能发展、安全与治理重大问题，不断推动人工智能治理领域的双边、多边合作。

2023年10月18日，习近平主席在第三届“一带一路”国际合作高峰论坛上提出《全球人工智能治理倡议》，这是中方积极践行人类命运共同体理念，落实全球发展倡议、全球安全倡议、全球文明倡议的具体行动。中方坚持“以人为本”理念和“智能向善”宗旨，坚持广泛参与、协商一致、循序渐进原则，积极支持以人工智能助力可持续发展，应对气候变化、生物多样性保护等全球性挑战，致力于弥合各国间“智能鸿沟”，推动各国人工智能技术平衡发展和成果共享，彰显了中国负责任的大国担当。

2024年4月，中非互联网发展与合作论坛发布《关于中非人工智能合作的主席声明》，划定“加强政策对话与沟通”“推动技术研发与应用”“促进产业合作与发展”“开展人才交流与能力建设”“筑牢网络与数据安全”等五个方面的合作要点，为中非加强人工智能领域合作，携手构建更加紧密的中非网络空间命运共同体，促进中非共同发展、共同繁荣描绘了具体蓝图。

2024年5月，中国和法国发布《关于人工智能和全球治理的联合声明》，双方共同强调人工智能必须为公共利益服务，各国开发和使用人工智能必须符合《联合国宪章》的宗旨和原则。双方强调加强国际合作，以弥合数字鸿沟，并提高发展中国家的人工智能能力，承诺充分致力于促进安全、可靠和可信的人工智能系统，坚持“智能向善（AI for good）”的宗旨，将依托联合国层面开展的工作，致力于加强人工智能治理的国际合作以及各人工智能治理框架和倡议之间的互操作性。

2024年7月，第78届联合国大会协商一致通过中国主提的加强人工智能能力建设国际合作决议，140多国参加决议联署。该决议强调人工智能发展应坚持以人为本、智能向善、造福人类的原则，鼓励通过国际合作和实际行动帮助各国特别是发展中国家加强人工智能能力建设，增强发展中国家在人工智能全球治理中的代表性和发言权，倡导开放、公平、非歧视的商业环境，支持联合国在国际合作中发挥中心作用，实现人工智

能包容普惠可持续发展，助力实现联合国 2030 年可持续发展议程。决议充分彰显出中国对人工智能发展和治理的负责任态度和重要引领作用。

三、积极构建全球人工智能共同治理体系的建议

全球人工智能技术快速发展，对经济社会发展和人类文明进步产生深远影响，给世界带来巨大机遇。与此同时，人工智能技术也带来难以预知的各种风险和复杂挑战。人工智能治理攸关全人类命运^①。当前，产业发展尚在初期，技术演进方向、应用场景落地都有不确定性，人类对于人工智能工作原理、风险机制的认知尚存在局限性。同时，人工智能治理又呈现跨国界属性，各国固然可以根据本国国情进行国内治理，但当人工智能所产生的影响对象是全人类时，一国国内治理势必产生外溢效应，需通过国际对话与合作，形成全人类对于推动人工智能发展和保障人工智能安全的基本共识。

为此，我们建议：应以人工智能健康与可持续发展为出发点，以基于发展的治理为准则，打破国家之间、机构之间、学科之间壁垒，共同分享最佳实践、主动对接国际标准、提升联通水平与互操作性，求同存异、相互尊重，致力于构建公正有效的人工智能国际治理体系，确保全球人工智能人才、资本、技术要素的自由流通以及公平、非歧视的市场准入，确保人工智能发展符合联合国可持续发展目标，造福人类命运共同体。

（一）求同存异，持续沉淀基于共识的全球人工智能治理理念

第一，构建公平公正、综合平衡的人工智能国际治理体系

首先，人工智能涉及全人类，应由全人类公平公正参与。各方应以负责任态度对待人工智能治理，看重治理规则公平性，确保人工智能治理权力平等、机会平等、规则平等，增强发展中国家代表性，开展面向发展中国家的国际合作与援助，不断弥合智能鸿沟和治理能力差距^②，在尊重各国政策和实践差异的基础上，推动人工智能治理形成广泛国际共识。

其次，发展与安全作为治理的一体两翼，理应综合平衡。应致力于在联合国框架内

① 《全球人工智能治理倡议》

② 《全球人工智能治理倡议》

团结全球更多元力量参与治理，不仅参与人工智能安全治理进程，同时参与推动技术开源开放、人才交流互通，让市场规则在算力、数据、算法的全球配置中发挥基础作用，让算力、模型、应用的飞轮演化出缩小全球智能鸿沟的自然力量，以产业均衡促进治理协同，以共同发展、共同繁荣维护共同安全、共生利益，确保人工智能可持续发展。

第二，构建有助于实现负责任地使用安全、可靠和可信的人工智能系统的国际治理体系

全球各方应共同防范、应对人工智能技术可能面临的失控、恶用和滥用风险，应始终倡导负责任地使用人工智能，反对将人工智能技术用于网络攻击、生成虚假信息等不良目的；应鼓励各方搭建并实施人工智能风险管理框架，建立和完善人工智能应用风险防范措施和跟踪机制，建立并畅通人工智能风险国际沟通机制和应急协调机制，通过可信任、全面、包容的对话，挖掘人工智能潜力，降低人工智能风险，确保人工智能技术的安全、可靠和可信。

第三，构建符合发展规律、有利产业繁荣的人工智能国际治理体系

首先，发展治理侧，对技术应用保持耐心，更多由市场力量推动。无论在美、欧还是中国，当前全球范围内生成式人工智能技术的成熟应用场景，都还有待进一步丰富。对技术的应用应当保持足够耐心，要防止“预设场景”导致行业浮躁。发展型治理中，政府力量应主要作用于基础研究、教育改革、人才政策优化、创新环境培育、市场预期引导等方面，同时强化企业创新主体地位，生产要素的汇集、配置，应充分发挥市场的主导作用，算力、数据问题从长远看均需依赖于市场的力量来解决。

其次，安全治理侧，对技术进步保持耐心，防止过度超前的监管措施压制行业发展预期。目前多数场景下，新旧人工智能技术混用是常态，而人类对于技术短期的过度乐观，往往会滋生“泡沫”、走向“超前宣传”，反过来可能在治理侧催生将“可能存在的风险”视作“已经来临”，进而导致过度监管、提前监管。对于传统人工智能，既有监管手段、工具相对成熟，而对于新型人工智能技术，由于对新技术的工作原理、风险机理，社会认知尚处于初步阶段，宜采取保持适度弹性与韧性的敏捷灵活的治理框架。

（二）承继创新，着力构建开放包容的全球人工智能治理体系

人工智能的全球治理应当是一个包容性的体系，政府、学术界、产业界、民间社会

和公众等利益相关方，都应有机会共同塑造人工智能的未来方向。

第一，着力构建、丰富完善各类多双边国际治理机制。首先，充分尊重并用好各类多边治理机制和既有原则框架。应助力联合国推出相关国际公约与标准规范，建立双边、多边等更多维度的政府间合作机制。一段时间以来，经济合作与发展组织（OECD）、联合国教科文组织（UNESCO）、世界卫生组织（WHO）、国际电信联盟（ITU）、世界经济论坛（WEF）等国际组织持续探索并先后提出多个人工智能的共识原则与发展规范。在生成式人工智能浪潮兴起背景下，相关原则、范式持续迭代更新，如联合国发布《全球数字契约》政策简报，就人工智能和其他新兴技术的敏捷治理提出四项目标，UNESCO 发布《生成式人工智能在教育和研究中的应用指南》，联合国人工智能高级别咨询机构发布《以人为本的人工智能治理》（中期报告）等等。其次，高度重视双边对话和相关治理机制构建。中国、美国、欧盟、加拿大、英国、日本、韩国、新加坡等国家，都提出了人工智能的治理框架或立法规范，其中一些国家也已开启了双边对话。以中美为例，作为人工智能发展最为迅速的两个国家，两国根据元首共识开启人工智能政府间对话，而智库、学界之间人工智能对话也蓬勃开展。同时，从去年的英国布莱切利峰会到今年的韩国首尔人工智能峰会，各方关于人工智能安全治理紧迫度的认识也在不断上升。

为此，我们建议：应平衡好存量与新增治理机构之间的关系，确保最大多数利益相关方能够公平参与治理，且有足够的资源来确保治理落地有效。人工智能理应为公共利益服务，各国开发和使用人工智能须符合《联合国宪章》的宗旨和原则^①。我们倡导首先在联合国框架下建立不同利益相关方都能有效参与的开放性对话平台和包容性治理机制，推动人工智能全球治理取得广泛共识，以此作为各类双边对话和治理合作的基础与前提。

第二，相信科学、依靠科学，以科学共同体力量助力国际治理机制完善。人工智能技术发展和安全防范事关全人类共同命运，需要凝聚全人类最顶尖学者的共同智慧。高性价比算力、高质量数据集、高端算法人才供给，最终都离不开科学理论与实践的持续进步。例如，大模型训练及推理势必加剧算力紧缺，但得益于包括阿里云在内的全球主

^① 《中华人民共和国和法兰西共和国关于人工智能和全球治理的联合声明》

要头部云企所提供的高性能、稳定和普惠的并行计算训练算力支持，以及高性价比、高弹性、高可用和就近合理分布的推理算力支持，全球主流模型服务价格一降再降，为人工智能大规模产业化应用奠定了基础，为全球算力资源的公正普惠分配找到了可行路径。此外，人类对于前沿人工智能的工作机理和风险研判，都尚处在初步阶段，技术演进带来的问题，终将依靠技术进步和产业发展而得到解决。降低模型幻觉、打击深度伪造、防止 AI 赋能网络攻击等“AI+”风险，都离不开 AI 安全技术与工程化能力的持续提升。深度学习与规模定律（Scaling Law）的不可解释性，只能依靠科学研究的进一步发展才可能寻得答案。一些头部的技术和治理专家尝试为 AI 系统划出红线，包括“系统开始产生欺骗行为”等，类似红线的类型、范围、执行都需要科学的持续进步来予以回应和落实。

为此，我们建议：在联合国人工智能高级别咨询机构基础上成立类似联合国政府间气候变化专门委员会（IPCC）的专门委员会和工作组，定期召开会议，开展专门研究，发布相关报告，以为国际治理提供科学可靠、具体可实施的智力支撑；其次，加强国际学术交流与合作，特别是推动开展人工智能基础研究领域的合作，在理论上更好地解释人工智能风险根源、预测人工智能发展潜力和制定人工智能发展规划。

第三，基于现实需求，以鲜活的产业实践为国际治理提供支撑。生成式人工智能浪潮新近袭来，作为一项新兴技术，由于其快速演进和不确定性，只有第一线的开发和应用者才能直接感知到其风险问题和治理关键^①。在长期实践中，产业界普遍通过提出准则指引、成立治理机构、开发治理工具等方式，践行负责任的人工智能。准则指引方面，微软、谷歌、阿里、腾讯、商汤等企业推出企业人工智能原则，普遍包括对社会有益、安全可控、保护隐私、公平、透明可解释等元素，如谷歌 AI 原则由 7 条正面清单和 4 条负面清单组成，正面清单包括对社会有益、避免加强不公平歧视、基于安全基础构建并增强安全性、与人类社会相适配并受到人的控制、注重隐私设计、服务于高标准科学实践、适用场景原则。治理机构方面，如中国早在 2017 年就发起成立人工智能产业发展联盟，中国信通院 2021 年发布《可信人工智能白皮书》，面对生成式人工智能浪潮，中

^①斯坦福《人工智能指数报告》显示，2023 年新发布的基础模型中，108 个来自产业界，仅 28 个来自学术界，其背后根源在于，传统算力供给模式无法适应新的 AI 研究范式：据统计，斯坦福 Sherlock 仅有 700 多张 GPU 卡，哈佛大学肯普纳自然与人工智能研究所仅有 600 张 GPU 卡，相当于微软当前最大单一 GPU 集群的 3% 左右。学术机构单一智算集群规模小、算力高度碎片化、算力供给缺乏灵活性和可扩展性，难以满足大项目对大算力的需求。

国政法大学、中国社科院面向全球发布了两个版本的人工智能立法专家建议稿。此外，科技企业内部普遍内设伦理委员会，定期输出本企业有关人工智能发展与治理的实践与思考。面对生成式人工智能浪潮，阿里巴巴去年发布《生成式人工智能治理与实践白皮书》，提出“发展多主体协同敏捷治理体系，构建全生命周期风险分类治理框架”的理念。治理工具方面，生成式人工智能浪潮以来，行业迭代既有治理工具，普遍围绕个人信息合规、内容安全保障、模型安全防控、知识产权保护探索等方面，在模型训练、服务上线、内容生成、内容传播四阶段形成了一整套新的工具体系，促进科技伦理治理原则技术化、工程化。如数据隐私方面，苹果公司发布其个人智能化系统“Apple Intelligence”，采取端、云两侧模型综合部署，同时为模型上云提出了完整的用户数据安全方案 PCC（Private Compute Cloud）。再如内容生成与传播方面，Meta 开源数字水印 Stable Signature，将数字水印嵌入 AI 自动生成的图片中，增强生成式人工智能安全。

为此，我们建议：在发展和安全两侧，行业实践应被充分理解和挖掘。应致力于构建更加包容、开放的技术开源社区，鼓励行业从业者相互分享并借鉴各国在人工智能不同领域的应用实践，共同探寻企业在人工智能发展与安全平衡过程中应有的角色和作用，为人工智能国际治理注入鲜活动力。

（三）循序渐进，分步打造综合系统的全球人工智能治理框架

首先，充分认识到人工智能治理问题的复杂性，做好四个统筹。

一是统筹发展与安全，充分发挥人工智能在帮助应对人类重大挑战、助力实现人类可持续发展方面的积极作用。二是加强全球伦理、立法、标准、测评等核心治理板块间统筹，完善涵盖“战略-产业-制度-工具”的一体化治理方案。三是加强国内治理与国际治理间统筹，完善一国国内主管部门间协同，构建国际间治理合作对接的“单一窗口”。四是运用既有治理经验和治理资源，与创新解决人工智能发展与治理发展中实际面临的痛点问题之间的统筹。

其次，尊重技术规律和产业实际，建设合作共赢的全球人工智能生态。

应从发展型治理与安全型治理两个维度，共建共享人工智能全球治理公共基础设施，如打造算力平台、数据平台、开源社区平台和治理工具平台、评测平台、风险信息互通和援助平台，推动人工智能的广泛应用和有效治理。

发展侧，应推动算力体系、数据要素和应用场景三位一体，着力加强算力和数据要素共享整合，降低重复建设成本，确保所有国家和地区都有机会参与人工智能的发展、享受技术红利。一方面，在全球范围内支持发展弹性普惠、全球共享的智能计算服务，以高性价比、高弹性、高可用和就近合理分布的算力支持，为 AI 规模化应用铺路；另一方面，打造多语种、跨文化、高质量的开放数据集，建设高效的大数据共享协作平台，提升清洗、标注等工程化能力。同时，推动全球人工智能开源社区发展，汇聚开源数据、模型、算力资源，促进技术交流互促。

安全侧，应保障不同人工智能治理框架和生态系统间的互操作性，倡导开放标准和开源技术，构建合作共赢的人工智能国际生态，鼓励各国开展数据跨境流动、评估结果互认、监管框架互通等议题的磋商，降低人工智能技术转移和应用的成本。当前，为全球所普遍接受的人工智能安全协调机制、标准设定、评测体系仍亟待建立，需要更为广泛的讨论和参与。

总言之，以生成式人工智能为代表的先进人工智能技术的演进、应用的落地、风险的防控、发展的治理，可能需要更长的周期，对此各个方面包括国际社会应保持一定耐心。未知的风险源自可能的发展，也只能靠发展来解决，不发展是最大的不安全和最大的风险。治理需要承继、更需要创新，一方面，需要基于市场和社会需求、积极推动技术和产业发展，另一方面，应当正视技术飞速发展与大范围应用带来的冲击与失衡问题，回应不同主体、不同国家各自关切，在交流对话中寻找最大公约数，就人工智能赋能发展与风险防范提供全球公共产品，在基于各国最佳实践的基础上探讨建立人工智能全球治理规则与治理机制，秉承“以人为本，智能向善”的理念，充分发挥人工智能在帮助应对人类重大挑战方面的积极作用，以人工智能健康发展支撑人类可持续发展未来。

作者：

顾登晨、刘明 阿里巴巴集团

盛舒洋、王净宇 清华大学人工智能国际治理研究院

审核：

肖茜、梁正 清华大学人工智能国际治理研究院

参考文献

- [1] 全球人工智能治理倡议[EB/OL]. 中央网信办, https://www.cac.gov.cn/202310/18/c_1699291032884978.htm, 2023.10.18.
- [2] Interim Report: Governing AI for Humanity [R], UN Secretary-General's AI Advisory Body, 2023.12.
- [3] Bengio Y, Hinton G, Yao A, Lan Xue, et al., Managing Extreme AI Risks Amid Rapid Progress [J], Science, 2024,384(6698): 842-845.
- [4] 薛澜,赵静. 人工智能国际治理: 基于技术特性与议题属性的分析 [J]. 国际经济评论, 2024, (3): 1-18.
- [5] 顾登晨. 静水流深: 美国人工智能治理的特征、趋势与启示 [EB/OL]. 阿里研究院, <https://www.secrss.com/articles/63209>, 2024.1.25.
- [6] 鲁传颖. 全球人工智能治理的目标、挑战与中国方案 [J]. 当代世界, 2024, (5): 25-31.
- [7] 高奇琦. 从大国协调到全球性机制: 人工智能大模型全球治理路径探析 [J]. 当代世界, 2024, (5): 12-18.
- [8] 贾开, 俞晗之, 薛澜. 人工智能全球治理新阶段的特征、赤字与改革方向 [J]. 国际论坛, 2024, 26 (3): 62-78+157-158.
- [9] 薛澜,贾开,赵静. 人工智能敏捷治理实践: 分类监管思路与政策工具箱构建 [J]. 中国行政管理, 2024, 40 (3): 99-110.
- [10] 薛澜. 新兴科技领域国际规则制定: 路径选择与参与策略 [J]. 人民论坛·学术前沿, 2023, (19): 15-21.
- [11] 梁正, 薛澜, 张辉, 曾雄. 人工智能治理框架与实施路径[M]. 北京: 中国科学技术出版社, 2023.
- [12] 梁正. 敏捷与协同: 人工智能治理理念与实践前沿[M]. 北京: 社会科学文献出版社, 2023.
- [13] 梁正. AI 治理大家谈[M], 北京: 人民邮电出版社, 2023.
- [14] 梁正, 何江. ChatGPT 意义影响、应用前景与治理挑战 [J]. 中国发展观察, 2023 (6): 15-19.
- [15] 庞祯敬, 薛澜, 梁正. 人工智能治理: 认知逻辑与范式超越 [J]. 科学学与科学技术管理, 2022 (9): 3-18
- [16] 姜李丹, 薛澜. 我国新一代人工智能治理的时代挑战与范式变革 [J]. 公共管理学报, 2022, 19 (02): 1-11+164.
- [17] 梁正, 张辉. 构建平衡包容的人工智能治理体系[J]. 中国发展观察, 2022 (12): 44-50
- [18] 刘志毅、梁正、郑烨婕, 黑镜与秩序: 数智化风险社会下的 AI 伦理[M], 北京: 清华大学出版社, 2022.
- [19] 傅莹. 人工智能的治理和国际机制的关键要素 [J]. 人民论坛, 2020, (04): 6-8.
- [20] 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 [J]. 国际观察, 2018, (4): 67-83.

◆ 全球人工智能治理与中国方案（征求意见稿）