



# 2021人工智能合作与治理 国际论坛报告

---

2021年12月

# 目录

一、论坛介绍	3
二、论坛嘉宾	3
三、论坛成果	7
四、论坛观点总结	9
(一) 开幕式讲话	9
(二) 主论坛 I 报告	11
(三) 主论坛 II 报告	13
(四) 主论坛 III 报告	15
(五) 专题论坛报告	18
五、2022 人工智能合作与治理国际论坛	28

## 一、论坛介绍

伴随新一轮科技革命和产业革命的深入发展，人工智能作为有望引领未来变革的战略性技术，在重塑生产方式、优化产业结构、提升生产效率、赋能千行百业的同时，也带来了安全、隐私、公平等问题和挑战。人工智能治理这一概念应运而生，推动构建平衡包容的人工智能治理体系已成为各方共同诉求。

由清华大学人工智能国际治理研究院（I-AIIG）主办，联合国开发计划署（UNDP）作为国际支持机构的第二届人工智能合作与治理国际论坛于 2021 年 12 月 4-5 日在清华大学成功召开。论坛以“如何构建一个平衡包容的人工智能治理体系”为主题，探讨构建适合人工智能健康发展的治理体系，分享人工智能技术前沿与治理经验，提出元宇宙未来治理的前瞻议题，推动建设人类命运共同体。

论坛形成了一系列重要观点与共识，受到国内外各界普遍关注与认可，超过 20 个国家的 70 余名专家学者、国际组织代表和企业代表出席，线上线下参会人数超过 500 人。海内外 40 余家媒体，60 多位记者参与，推出报道 160 余篇，直播环节收看人次累计达 300 万，微博相关话题传播超过 200 万次。

## 二、论坛嘉宾

### 中方嘉宾（按报告顺序）

**汪永清** 中共十九届中央委员，十三届全国政协副主席、党组成员

**邱勇** 清华大学校长、中国科学院院士

**杨斌** 清华大学副校长、人工智能国际治理研究院管委会主任

**姚期智** 图灵奖得主、中国科学院院士、美国国家科学院外籍院士、清华大学交叉信息研究院院长、人工智能国际治理研究院学术委员会主席

**李 萌** 科技部副部长

**徐晓兰** 工信部副部长

**马升琨** 外交部军控司副司长

**薛 澜** 清华大学文科资深教授、苏世民书院院长、人工智能国际治理研究院院长、国家新一代人工智能治理专业委员会主任

**龚 克** 世界工程组织联合会主席、中国新一代人工智能发展战略研究院执行院长、清华大学人工智能国际治理研究院学术委员

**苏 竣** 清华大学智能社会治理研究院院长、智库中心主任、人工智能国际治理研究院管理委员会副主任

**梁 正** 清华大学人工智能国际治理研究院副院长

**曾 毅** 中国科学院自动化研究所中英人工智能伦理与治理中心主任、清华大学人工智能国际治理研究院首席专家

**张希良** 清华大学能源环境经济研究所所长

**吴力波** 复旦大学大数据研究院副院长，能源经济与战略研究中心主任

**田 丰** 商汤智能产业研究院院长

**肖 茜** 战略与安全研究中心副主任、人工智能国际治理研究院副院长

**陈 琪** 清华大学战略与安全研究中心秘书长、社会科学学院国际关系学系教授

**李驰江** 中国军控与裁军协会副会长兼秘书长

**李睿深** 清华大学战略与安全研究中心客座研究员

**朱启超** 国防科技大学国防科技战略研究智库副主任、研究员

**于 洋** 清华大学交叉信息研究院助理教授

**陈 玲** 清华大学产业发展与环境治理研究中心主任

**陈楸帆** 科幻作家

**孙茂松** 清华大学计算机科学与技术系长聘教授、人工智能研究院常务副院长、人工智能国际治理研究院管理委员会委员

**张 钹** 中国科学院院士、清华大学人工智能研究院名誉院长、人工智能国际治理研究院学术委员

**高 文** 中国工程院院士、北京大学信息科学技术学院院长、鹏城实验室主任、清华大学人工智能国际治理研究院学术委员

**沈向洋** 美国国家工程院外籍院士、英国皇家工程院外籍院士、香港中文大学（深

圳）校长讲座教授、清华大学高等研究院双聘教授、前微软全球执行副总裁

**张亚勤** 中国工程院外籍院士、清华大学智能产业研究院院长、人工智能国际治理研究学术委员

**马维英** 清华大学惠妍讲席教授、智能产业研究院首席科学家

**夏华夏** 美团副总裁、首席科学家

**秦 尧** 华为公司人工智能战略与产业发展副总裁

**陶大程** 京东探索研究院院长、澳大利亚科学院院士

**宋继强** 英特尔研究院副总裁、英特尔中国研究院院长

**黄晓庆** 达闼创始人、董事长兼 CEO

**徐 健** 地平线首席生态官、副总裁

**申卫星** 清华大学法学院院长、智能法治研究院院长

**孟庆国** 清华大学中国电子数据治理工程研究院院长、互联网治理研究中心主任

**段伟文** 中国社科院哲学所科技哲学研究室主任、研究员，科学技术和社会研究中心主任

**沈 阳** 清华大学新闻与传播学院新媒体研究中心执行主任

**韦 青** 微软（中国）CTO

**余 晨** 易宝支付联合创始人兼总裁

**籽 岷** Bilibili up 主

**何 展** 英伟达中国区 Omniverse 业务发展经理

**宋 婷** AI 和区块链艺术家

**朱旭峰** 清华大学公共管理学院教授、常务副院长，全球可持续发展研究院执行院长、科技发展与治理研究中心副主任

**贾晓丰** 北京市大数据中心数据管理部负责人

**外方嘉宾（按报告顺序）**

**白雅婷** 联合国开发计划署驻华代表

**玛丽亚－弗朗切斯卡·斯帕托利萨诺** 联合国助理秘书长、秘书长技术事务特使办公室主管官员

**徐浩良** 联合国助理秘书长、开发计划署政策与方案支助局局长

**温德尔·瓦拉赫** 卡内基理事会资深研究员、清华大学人工智能国际治理研究院学术委员

**罗欣顿·麦德拉** 加拿大国际治理创新中心总裁

**维奥兰特·迪卡诺萨** 联合国开发计划署中国代表处首席经济学家

**尤金尼奥·巴尔加斯·加西亚** 巴西总领事馆副总领事兼科学、技术和创新主管，

联合国大会和平与安全、人道主义和法律事务高级顾问和团队负责人

**罗汉·萨马拉吉瓦** LIRNEasia 智库主席（斯里兰卡）

**瓦苏基·沙斯特里** 英国皇家国际事务研究所亚太项目副研究员

**南吉拉·桑布利** 卡内基国际和平研究院技术与国际事务项目研究员

**南南·伦丁** 瑞典驻华使馆科学创新处参赞

**肖恩·欧·海格缇** 剑桥大学生存风险研究中心联合主任、未来智能研究中心项目

主任

**艾玛·鲁特坎普·布洛姆** 联合国教科文组织人工智能伦理特设专家组主席

**文森特·穆勒** 埃因霍芬理工大学教授、英国艾伦·图灵研究所研究员

**阿曼迪普·辛格·吉尔** 联合国秘书长数字合作高级别小组秘书处前执行主任

**马克·芬德利** 新加坡管理大学法学教授、人工智能与数据治理中心主任

**丹妮特·加尔** 联合国秘书长办公室前 AI 主任、剑桥大学未来智能研究中心外籍

研究员

**张薇** 联合国开发计划署助理驻华代表

**费利克斯·克鲁齐格** 墨卡托全球公域和气候变化研究所教授

**西尔文·杜兰顿** BCG 董事总经理，全球资深合伙人

**普里亚·唐蒂** 气候变化人工智能联合创始人兼总裁

**吕晶华** 人道主义对话中心东北亚项目主任

**瓦季姆·科久林** 俄罗斯政策研究中心、新兴技术和全球安全项目、亚洲安全项目

主任

**佐藤丙午** 日本拓殖大学国际学部教授、海外事务研究所教授兼副所长

**让-马可·瑞克利** 日内瓦安全政策中心全球和新兴风险部主管

**埃里克·里查德森** 人道主义对话中心资深顾问

**R.S. 潘瓦尔** 未来战争分析师、印度陆军信号兵团第 57 任中校（退役）

**巴里·奥沙利文** 科克大学计算机科学与信息技术学院教授

**保罗·沙瑞尔** 新美国安全中心副主席兼研究主任

**伊德·加尼** 亨氏信息系统与公共政策学院机器学习系数据科学与公共政策中心主任

**丹尼尔·弗洛伊德** 麻省理工学院斯隆管理学院运营管理助理教授

**毕文韬** 联合国妇女署项目和伙伴关系专家

**苏珊·阿西** 美国国家科学院院士、美国艺术与科学院院士、约翰·贝茨·克拉克

奖得主、斯坦福大学商学院技术经济学教授、以人为本人工智能研究所副所长

**约兰达·吉尔** 国际人工智能协会前主席、清华大学人工智能国际治理研究院学术

委员

**克里斯托夫·吕特格** 慕尼黑工业大学人工智能伦理研究所主任、清华大学人工智

能国际治理研究院学术委员

**德地立人** 清华大学产业发展与环境治理研究中心执行理事、原中信证券股份有限公司

董事总经理

**亚历山大·克里比茨** 慕尼黑工业大学人工智能伦理研究所研究人员

## 三、论坛成果

### （一）形成对人工智能治理国际合作的基本共识

第一，需在人工智能治理等多领域秉承合作共赢的理念。人工智能治理已成为一项重要的全球议题，国际社会必须携起手来、共同应对人工智能发展带来的风险挑战，加强各国间人工智能制度政策的沟通，积极开展旨在应对新挑战的前瞻性对话。

第二，需建立正式的、常态化的、多主体参与的人工智能治理国际对话平台，定期讨论、共同研究、多方采纳。联合国在为各方达成协议、取得共识等方面起到了非常大的作用，人工智能治理国际对话平台应在联合国的框架下，汇聚跨国界、跨代际、跨领域和跨学科的合作来制定必要的人工智能全球治理框架。

第三，需在人工智能伦理与治理中建立起跨文化信任。首先，应进行关于文化概念的全球合作。其次，通过跨文化合作了解不同文化对人工智能的使用和监管，使其跨越地缘政治边界。最后，应把握好普世性与差异化的文化伦理平衡点，推动以互相尊重、加强信任、包容对话为基础的人工智能治理跨文化国际合作。

第四，要推动共识性的人工智能伦理标准和治理规则的建立。在此过程中，要推动协商互鉴，建立起跨文化信任，尊重和包容差异，统筹发展与安全，兼顾效率

与公平，防止治理规则政策取向顾此失彼、避重就轻和歧视性排他性，共建具有广泛共识的全球人工智能治理规则和伦理标准的平衡性框架。

第五，需要积极探索双边、多边合作机制。充分发挥中美两国以外第三方国家的作用，共同建构包容性的人工智能治理模式。

## （二）推动构建平衡包容的人工智能治理体系

第一，提出人工智能治理体系基础框架。在价值导向方面，要以“底线思维、人本思维、发展思维、全球思维”为价值导向。在治理主体方面，需要“政府、企业、公众、大学与科研机构、媒体、非盈利组织、国际组织”等多主体参与，多元治理主体在人工智能社会技术系统中拥有不同的权威、资源、利益与限制，通过各种正式与非正式渠道进行广泛互动，构成治理机制复合体。在治理对象方面，一是数据，需要关注数据自身的安全，数据自主可控和宏观安全；二是算法，要聚焦算法的稳定性、安全性、可解释性、公平性；三是算力，需要寻求多边合作共赢，避免技术垄断；四是外部环境，如何应对人工智能技术的外部性，尤其是负外部性影响，给出必要的补偿或者救济；五是场景，需要考虑人工智能应用的各种具体场景。

第二，提出多路径多层次的人工智能治理机制。一是形成基本价值共识。从目前全球范围内发布的各种准则来看，虽然各具特色，但最核心的内容还是从包容、共享、审慎、负责的价值原则对人工智能发展提供相关引导；二是促进治理主体分工协作。大致上可识别出四类主体：政府部门、人工智能技术提供者、人工智能技术使用者、社会公众和第三方组织。政府部门要对人工智能技术提供者（包括研发组织、平台组织、高等院校等）进行一定程度的赋权，同时保持有效的监督。人工智能技术提供者要积极和政府部门进行有效互动，充分赋能政府部门，提高公共服务的质量和效率。人工智能技术使用者要与技术提供商进行合作和互动，同时要接受政府部门的监管。社会公众、第三方组织等社会主体要积极参与，对全要素、全对象和全过程提供及时有效的监督；三是保持治理机制迭代优化。传统技术治理领域，技术应用往往需要经过几十年的扩散过程，在社会上只能达到70%、80%的渗透率。但是，人工智能技术的创新发展和应用扩散非常快，技术发展和治理规则之间存在时间差。面对这种情况，要保持各种治理机制和治理模式相结合。首先是探索性治理或是前瞻性治理，预想到潜在的风险并做好风险预案。其次是适应性治理或者敏捷治理，要有效地平衡治理效果和治理节奏。

第三，应用多样化的治理工具组合。应用“宏观-中观-微观”多维治理工具，

构建人工智能综合治理框架体系，推动人工智能健康发展。

## （三）提出元宇宙未来治理前瞻议题

第一，元宇宙应秉承以人为本。开发虚拟世界，旨在用其帮助、强化现实世界，而非替代。

第二，对元宇宙的治理应有前瞻性的考量。元宇宙的构想与构建具有超越性，如果继续以信息和数据作为构建世界的关键路径，将会是整合多种新技术而产生的新型虚实相融的互联网应用和社会形态。未来，对元宇宙的监管和治理可能面临的不是单一维度的技术体系——唯一的大写的元宇宙帝国，而是一个多维整合的技术体系——多样化的元宇宙共和国或多种微世界的联合体。有必要重新审视目前分散但有联系的各个虚拟空间对现实世界的冲击。

第三，元宇宙未来治理需要共治。当前，元宇宙整体发展处于初级阶段，其技术应用还处于低级层次。对于元宇宙的治理，可以采取边调研边发展边治理，确保不出现重大风险。未来，元宇宙治理需建立在充分调研的基础上，经过多元主体的社会大讨论，实现发展与治理的平衡，避免一刀切式的治理框架。

## 四、论坛观点总结

### （一）开幕式讲话

中共十九届中央委员，十三届全国政协副主席、党组成员汪永清表示，中国人工智能治理坚持以人民为中心、坚持增进人类福祉、坚持鼓励科技创新、坚持统筹发展与安全、坚守公平正义理念，契合联合国和G20有关人工智能治理精神，目的是促进人的全面进步、促进经济社会生态可持续发展。中国政府一向主张加强人工智能发展与治理国际交流合作，倡导共商、共建、共享解决全球数字治理问题，携手打造开放、公平、公正、非歧视的数字发展环境。面对人工智能发展带来的风险挑战，国际社会必须携起手来、共同应对，不断提升人工智能全球治理水平。要推动开放包容，秉持人类命运共同体理念，摒弃强权霸道逻辑，摆脱意识形态偏见，最大限度地增强合作机制、理念、政策的开放性和包容性，构建人工智能全球治理的负责任格局。要推动协商互鉴，坚守和平、发展、公平、正义、民主、自由的全人类共同价值，建立起跨文化信任，尊重和包容差异，统筹发展与安全，兼顾效率与公平，防止治理规则政策取向顾此失彼、避重就轻和歧视性排他性，共建具有广



泛共识的全球人工智能治理规则和伦理标准的平衡性框架。要推动共治共享，创建人工智能治理各方参与的开放性平台，通过跨学科、跨领域、跨地区、跨国界的交流合作，推动相关各方遵循基本科技伦理，把进步的价值取向嵌入人工智能数理模型，形成人工智能向善的力量，推动形成人工智能治理的最佳实践，形成各方积极参与的共治共享良好局面。

清华大学校长、中国科学院院士邱勇表示，人工智能是第四次工业革命的标志性技术。近年来，智能化正在深刻改变人们的生产生活方式，智能时代的“大门”已经缓缓开启。然而，在提高社会生产力、赋能治理现代化的同时，人工智能也带来了公共安全风险、隐私侵权、数字鸿沟、责任困境、伦理失范等许多新的挑战。人工智能治理必须协同多元主体、兼顾多维目标、融合多元价值，加强全球合作与对话，构建平衡包容的人工智能治理体系。大学作为人类科技文明的重要策源地，不仅要在人工智能基础理论与技术前沿努力突破创新，也要践行科技向善的文化理念，塑造良善的人工智能价值与伦理，推动建立科学合理的人工智能国际治理体系。

联合国开发计划署驻华代表白雅婷指出，人工智能在疫情应对、抗击气候变化中都发挥着重要作用。与此同时，人工智能也存在算法偏见和侵犯个人隐私等风险。随着人工智能应用在世界各地愈加普及，需要就公平道德的人工智能治理体系的构建路径达成共识，国际合作至关重要。因此，要确保围绕人工智能开展的对话真正具有全球性，让每个国家都有机会发声。

科技部副部长李萌表示，中国秉持科技向善，高度重视人工智能治理。近年来，科技部围绕推动人工智能治理先后发布实施了《新一代人工智能治理原则》《新一代人工智能伦理规范》，重点开展了建立统筹推进机制、启动实施社会实验、发挥企业和地方作用、积极参与国际治理合作等工作。未来，人工智能治理要倡导敏捷治理理念，实现人工智能发展与安全的平衡；倡导包容共享理念，实现不同群体共享智能红利的平衡；倡导可持续发展理念，实现智能化与绿色化的平衡；倡导人类命运共同体理念，实现差异发展与合作共赢的平衡。

工信部副部长徐晓兰表示，在“互联网+人工智能”的驱动下，人类的生活方式发生着深刻变革；在“工业互联网+人工智能”的驱动下，产业的生产方式将发生深刻变革，人工智能还将深刻影响社会治理的现代化进程。中国高度重视人工智能产业发展。工业和信息化部认真贯彻落实党中央国务院决策部署，组织开展了人工智能重点任务揭榜工作，组建了智能传感器、智能网联汽车等国家制造业创

新中心，推动北京等8个人工智能创新应用先导区建设等。未来，工信部将携手相关各方，在共促人工智能持续健康发展、共建包容共享的人工智能治理规则、共创国际合作新局面等方面大力合作，为推动人工智能国际治理体系建设贡献中国智慧，并为国际科技交流与合作增添更多务实成果。

外交部军控司副司长马升琨表示，中国高度重视人工智能领域安全治理，自始至终秉持负责任态度，统筹推进人工智能技术与风险防范，及时出台研发人员伦理准则和规范，明确监督机制和要求。围绕未来如何进一步规范人工智能军事应用问题，建议如下：从国际法层面，应遵守国际法和国际关系基本准则；从道德层面，应秉持“智能向善”理念；从技术层面，应加强对人工智能技术的评估和监管；从治理层面，应积极倡导全球共同治理。

联合国助理秘书长、秘书长技术事务特使办公室主管官员丽亚-弗朗切斯卡·斯帕托利萨诺指出，人工智能是联合国秘书长发布的《数字合作路线图》中提到的关键数字挑战之一。作为一项能够带来真正变革的技术，人工智能的应用也伴随着严峻挑战。在充分发挥人工智能益处的同时，需要在全世界范围内进行更好的引导与合作。联合国秘书长在《我们的共同议程》报告中强调了要加强人工智能领域的全球监管，提议要建立一个由多利益攸关方参与的全球人工智能合作新机构，特别强调要纳入全球南方国家及其观点的重要性。各国必须共同努力建立普惠包容、反应迅速、且行之有效的全球合作架构，以此应对人工智能带来的各项挑战。

图灵奖得主、中国科学院院士、美国国家科学院外籍院士、清华大学交叉信息研究院院长、人工智能国际治理研究院学术委员会主席姚期智指出，随着人工智能与经济社会的相互嵌入，保持科学技术与经济社会发展之间的良性互动与协同越来越重要，这就需要我们做好人工智能技术开发、应用与治理之间的平衡。然而，人工智能对数字经济时代的隐私保护和数据安全也提出了挑战，不断精进的算法、高效的算力和隐私保护是人工智能理论的新方向。“平衡包容”的人工智能治理体系的建构，将进一步释放人工智能的潜能，更大程度的保证人工智能的安全性。

## （二）主论坛 I 报告

主论坛 I 围绕“如何构建一个平衡包容的人工智能治理体系”这一主题展开。

清华大学文科资深教授、苏世民书院院长、人工智能国际治理研究院院长、国家新一代人工智能治理专业委员会主任薛澜指出，当前人工智能治理已从理念探讨走到了实践探索的前沿，人工智能治理在数据、算法、算力、场景等层面面临挑战，

未来人工智能治理需建构“政府、企业、公众、大学与科研机构、媒体、非盈利组织、国际组织”等多主体参与，以“底线思维、人本思维、发展思维、全球思维”为价值导向，以“算法、数据、算力、外部环境、场景”为治理对象，应用“宏观—中观—微观”多维治理工具的人工智能综合治理框架体系，通过“形成价值共识、多主体分工协作、治理理念迭代优化”等治理机制推动人工智能健康发展，促进人类福祉提升。

世界工程组织联合会主席、中国新一代人工智能发展战略研究院执行院长、清华大学人工智能国际治理研究院学术委员龚克认为，人工智能治理体系首先应是基于维护人权和实现可持续发展的体系。其次，应是以促进而非遏制人工智能创新发展为目标的体系。第三，应是多利益攸关方共治的开放体系。第四，应是综合运用技术、法规、教育多种工具的体系。第五，应是融入人工智能平台的有效体系。人工智能是一种新兴生产力，任何东西都无法阻挡新兴生产力的发展。人工智能是一种新兴生产力，任何东西都无法阻挡新兴生产力的发展。因此，人工智能治理应该促进其创新以造福人类和地球，而不是通过将其限制在现有体系中来遏制其发展。

联合国助理秘书长、联合国开发计划署政策与方案支助局局长徐浩良指出，人工智能已在世界各地推动落实可持续发展目标，也推动了教育、交通等其他领域的发展，人工智能的发展与治理需要并肩而行。虽然人工智能有助于扩大规划项目的规模，但也由于知识匮乏、疏忽或因恶意，加重了偏见、排斥和歧视。因此，需要跨国界、跨行业和跨代际的合作来制定必要的治理框架，比如能让政府、国际组织、科技公司和民间团体共同参与的多利益攸关方模式，这样有助于建立信任，识别偏见、歧视等风险，并加快项目实施进度。

卡内基理事会资深研究员、清华大学人工智能国际治理研究院学术委员温德尔·瓦拉赫认为，21世纪需要一个新的国际合作方式来治理人工智能。人工智能国际治理机制应该具有敏捷性、适应性、预见性、响应性和包容性，中美两国参与制定人工智能国际治理方案，将使其更具有可行性。每个国家需要处理好对国家安全和国际合作双重需求之间的关系，接受国际标准的同时，积极开展旨在应对新挑战的前瞻性对话。全球正在大量制定并实施人工智能治理的标准，各国之间的标准也会产生冲突，应尊重文化差异，并在经过多个阶段的发展之后，形成有效的人工智能国际治理机制。

清华大学智库中心主任、智能社会治理研究院院长、人工智能国际治理研究院

管理委员会副主任苏竣认为以人工智能等为代表的新兴科学技术日新月异，给世界各国的经济发展、社会治理和人民生活带来重大而深远的影响。智能社会的风险包括技术风险、市场风险、社会风险、认知风险，人工智能治理需构建有人文温度的智能社会治理体系，这需要用科学的方法研究应对科学技术带来的问题。复归“实验主义治理”这一传统的公共管理的思潮，遵循科学、规范、量化、循证的研究范式，充分利用我国人工智能应用广泛、场景多样的优势，积极开展人工智能社会实验，探索智能社会治理的中国道路。

加拿大国际治理创新中心总裁罗欣顿·麦德拉表示，人工智能从数据输入到技术输出的治理至关重要。数据层面，一是要建立全球共同接纳的伦理标准，将其作为人工智能的基础价值观；二是要建立数据使用过程中的全球普遍性标准，保证数据安全；三是要建立算法问责机制，保障数据被善用。技术层面，一是要构建更加高效的全球知识产权体系，完善数据共享和知识产权机制；二是要建立全球性的数据库，用新的范式驱动创新。未来，一方面需要建立包括政府、私营部门、非政府组织等共同参与的人工智能国际合作机制，充分发挥中美两国以外第三方国家的作用。另一方面，需要建立健全“以人为本”的全球人工智能工作委员会，传递正确的价值观。

就“如何推进人工智能领域的国际合作和治理”这一问题，上述专家一致认为应尽快建立一个国际性的人工智能对话平台。平台应在联合国框架下和已有的对话机制下进行设计，并汇集来自各个领域、不同专业的专家，推动对人工智能治理问题的全面研究。

### （三）主论坛 II 报告

主论坛 II 围绕“人工智能技术前沿与治理”这一主题展开。

美国国家科学院院士、美国艺术与科学院院士、约翰·贝茨·克拉克奖得主、斯坦福大学商学院技术经济学教授、以人为本人工智能研究所副所长苏珊·阿西认为，大学在指导人工智能创新方面可以发挥优先引导的关键作用。由于私营部门的技术人员缺乏伦理、哲学方面的训练，难以开发出具有可解释性的算法框架，深化这类研究能够在人工智能治理的问题识别、建立开发实践框架、提供指引等方面发挥重要作用。此外，由于数据可以带来巨大的规模效应，当前“软件即服务”的平台经济模式已非常普及。人工智能和数据需求可能带来“伪”市场集中，因此，未来对“机器换人”的预测非常具有挑战性，需要重新关注和思考人工智能如何用于

应对老龄化等公共管理问题，使基于人工智能的公共服务变得更加高效。

国际人工智能协会前主席、清华大学人工智能国际治理研究院学术委员约兰达·吉尔认为，由于人类对智能机制认知不足、智能行为本身的复杂性、观测手段的有限性以及个体知识、职业、信仰、文化背景等的差异性，导致当前人工智能研究中面临着一系列挑战，因此，需要加强人工智能基础研究工作，这需要跨领域、跨学科共同努力。当前，理解人工智能机理和构建人工智能世界模型是人工智能研究面临的两大挑战。一方面，理解人工智能机理需要构架“感知-思考-行动”的智能模型，加强对大脑思维机理的理解，建议借鉴神经科学研究联合体的有益经验，建立全球性的人工智能研究数据库，形成全球共享的研究社区。另一方面，构建人工智能世界模型则需要建立在人类经验、社会习俗、专业技能的基础上，建议建立类似于自由协作式的知识库，通过全民民众参与，推动知识在全球层面共享。

中国科学院院士、清华大学人工智能研究院名誉院长、人工智能国际治理研究院学术委员张钹指出，由于深度学习等算法存在不可解释性，导致前两代人工智能算法存在着公平性、安全性问题和不可靠、不可信等缺陷。发展第三代人工智能关键在于发展可解释的、鲁棒的人工智能理论和方法，开发安全、可信、可靠、可扩展的人工智能技术，以“数据驱动+知识驱动”构建支持可解释的人工智能算法的深度平台，赋能人工智能安全与防御优化。从数据中真正获取智能要靠知识的帮助与引导，并需要政策法规对数据使用的正确规范，充分利用知识、数据、算法和算力四个要素结合，推动人工智能的创新发展。

中国工程院院士、北京大学信息科学技术学院院长、鹏城实验室主任、清华大学人工智能国际治理研究院学术委员高文指出，当前人工智能发展处于新一代人工智能向强人工智能发展的关键阶段，至2030年，我国人工智能发展总体要达到世界领先水平。从战略问题看，中美欧三方在人工智能人才、研究、开发、应用、硬件、数据等方面竞争激烈，当前我国人工智能发展在战略政策、数据资源、应用场景、潜力人才方面具有优势，而在基础理论、原创算法、关键部件、国际平台、高级人才等方面还存在短板。从战术问题看，人工智能2.0需采用基于大数据的统计解决大规模人工智能应用需求，鼓励各种可能的强人工智能探索，“可解释机器学习+推理”和“仿生系统+AI大算力”是可能的技术路线图；在安全问题层面，强人工智能的安全风险主要来源于模型的不可解释性、算法和硬件的不可靠性和自主意识的不可控性，人工智能2.0应采用DPI与“防水堡技术”解决数据安全与

隐私保护，重视探索人工智能伦理问题，并基于“理论-技术研究-应用”的阶段性采取不同的风险防范策略。

美国国家工程院外籍院士、英国皇家工程院外籍院士、清华大学双聘教授、前微软全球执行副总裁沈向洋认为，人工智能已经应用于生活和工作的方方面面，目前甚至在法律上也具有一定的应用，比如美国已经有很多法庭用机器学习和人工智能方法帮助判刑，包括决定刑期这样非常重要的问题。但是我们还无法理解一些人工智能决策的缘由。未来发展过程中不能只看见人工智能决策的“黑箱”，应该打开“黑箱”，探究和理解其中的具体内容和因果关系，要做可解释性的人工智能。同时，负责任的人工智能应具备公平性、可靠性、隐私性、包容性、透明性和责任性的特点，作为新兴领域，还需要向其他领域学习，从而更好的服务于人类。

中国工程院外籍院士、清华大学智能产业研究院院长、人工智能国际治理研究院学术委员张亚勤认为，“碳中和”是人类能源结构的又一次变革。“碳中和”既是可持续发展的必然选择，又是产业结构调整 and 发展的重大机遇。企业在“碳中和”背景下都面临转型增效的压力。人工智能+物联网是智联网，智联网可以赋能绿色计算，助力“碳中和”。智联网助力“碳中和”主要包括三个环节：首先，由数据驱动和人工智能优化引擎来实现智能决策。其次，多参数全链系统配置优化。最后，通过多源多维异构感知融合实现智能感知。智联网可用于能源融合、降低ICT产业的碳排放和推动新兴产业发展等。

与会专家一致认为，需在人工智能治理等多领域秉承合作共赢的理念，定期讨论、共同研究、多方采纳。推动共识性的人工智能伦理标准和治理规则的建立，提升在算法规则、数据安全等方面的全球治理能力，共同建构包容性的人工智能治理模式。

#### （四）主论坛 III 报告

主论坛 III 围绕“元宇宙未来治理前瞻”这一主题展开。

中国社科院哲学所科技哲学研究室主任、研究员、科学技术和社会研究中心主任段伟文认为，人类目前为自己构建的技术世界建立在赛博物理空间(CPS)之上，不论是否称为元宇宙，下一步如果继续以信息和数据作为构建世界的关键路径，将会是整合多种新技术而产生的新型虚实相融的互联网应用和社会形态。亟待深入思考的是，元宇宙的构想与构建具有超越性，可能会把人类带进一个超历史空间。这个空间不仅仅涉及现实、虚拟及扩展现实等感知体验世界，更是一个超越现实与虚



拟体验，涉及符号/象征、想象等观念和意义创造的世界。如果考虑到符号/象征、想象这一更高的维度，元宇宙的构建可更多地依靠人们自主的意识和观念的创造与合适的技术实现方式的结合，而不一定要走向单纯依赖技术无限升级的加速主义，甚至落入单一趋同的模式。若认识到这一点，未来，对元宇宙的监管和治理可能面临的不是单一维度的技术体系——唯一的大写的元宇宙帝国，而是一个多维整合的技术体系——多样化的元宇宙共和国或多种微世界的联合体。对元宇宙的治理应有前瞻性的考量。

清华大学新闻与传播学院新媒体研究中心执行主任沈阳认为，元宇宙是虚实相融的新世界，元宇宙实现了对时空的拓展，其未来面临着诸多风险，如虚拟人的归属问题，虚拟人的责任困境，人机交互下个人认知异化和行为异化等。当前，元宇宙整体发展处于初级阶段，其技术应用还处于低级层次。对于元宇宙的治理，可以采取边调研边发展边治理，确保不出现重大风险，又能积极推进元宇宙产业发展。近期可在现有的制度、法律、政策框架内来讨论，将当下的治理抓手落在个人权力和利益关系层次。未来元宇宙治理需建立在充分调研的基础上，经过多元主体的社会大讨论，实现发展与治理的平衡，避免一刀切式的治理框架。要在发展过程中发现问题、分析问题、解决问题，实现精准式的动态治理。

中科院自动化所中英人工智能伦理与治理中心主任、清华大学人工智能国际治理研究院首席专家曾毅认为，元宇宙相关的若干概念目前在科学上十分不清晰并有误导性。当“元宇宙”相关基本概念、愿景和应用仍然看似有重大风险时，尤其不应当将青少年推向这个未知的空间。目前，对这一概念的刻画还相当不扎实。其次，在通用人工智能还未到来之时，互联网、人工智能所存在的所有风险就足以使人类社会受到巨大冲击，由于“元宇宙”基于现代网络与人工智能技术发展，其风险将会在此上基础上叠加。最后，从文化背景看，虚拟和现实世界，都是历史产物。重塑虚拟社会，历史不可磨灭。在虚拟平台进行共治，文化互鉴和互信是重要的“基础设施”。从技术角度来看，当我们许诺给全球公众和政治系统概念的时候，在提出经得起科学推敲的基础理念前提下，要考虑技术可行性和社会应用合理性。完全脱离真实世界重塑的新概念会将人类带向真正的生存风险。

微软（中国）CTO 韦青认为，每个人对元宇宙有不同的理解。元宇宙应该以人为本，应该用来强化现实世界，而非替代物理世界，不能让人们沉迷于虚拟世界。越来越多的人发现，虚拟空间与真实空间都需要治理，都需要有秩序。但什么是好

的治理，目前仍没有明确的答案，需要人类社会一起达成共识。在数字社会发展过程中，数字鸿沟是一个很大的挑战。技术向前迈进的同时，不能落下一部分人。面对新技术，需要有足够的信心认为技术能够为人类的美好生活做出贡献。

易宝支付联合创始人兼总裁、《元宇宙通证》作者余晨认为，创造的虚拟世界都可以被视为元宇宙，虚拟并不可怕。纵观人类历史，就是一个由实入虚的过程。元宇宙及现代社会很多问题的根源在于很多人混淆了符号和其所指的对象、混淆了拟像与真实。虚拟空间需要秩序，也需要治理。对于新技术可能的威胁，不能与之前的技术进行简单的归纳和类比，毕竟人工智能技术可能不需要拥有意识就足以毁灭人类，人们往往低估了新技术的次生危害。原则上技术是中立的，但技术是人性的“放大器”，对新技术的治理应该是动态的过程。元宇宙给人类带来创造价值可能性的空间，人们需要将恶的部分排除，将善的部分留下。

英伟达中国区 Omniverse 负责人何展认为，元宇宙是一个概念，而非一项技术。由于近年来许多数字技术的出现、落地和爆发，元宇宙这一概念被广泛讨论。确切来讲，元宇宙是包括人工智能、机器视觉、物理仿真模拟、高性能计算等许多技术的融合，进而形成虚实交融的新业态。元宇宙之“元”，代表客观存在的世界与人的主观意识、意愿和谐共处，需要服务于真实世界。比如开发点餐机器人，帮助人们针对喜好推荐合适的食物，过去几十年，大多数人已经率先体验了虚拟世界带来的红利。但有趣的是，如果能够区分现实世界和元宇宙之间的不同，人们往往会选择从锻炼自身做起优化自己的现实生活，实现虚实之间的平衡。

Bilibili up 主籽岷认为，人们在元宇宙中的身份，不仅只是代表一个人的虚拟身份，而是一个集合。当前，许多人可能不清楚正在使用的就是元宇宙，或者根本不了解这一概念。对元宇宙概念的理解，主要是基于虚拟与现实技术的讨论。从虚拟世界和现实世界的关系来看，虚拟世界目前是一个雏形，而现实世界是已成型的。需要谨记的是，开发虚拟世界，旨在用其帮助现实世界，并且应多方合作，使用元宇宙来帮助需要帮助的人。

AI 和区块链艺术家宋婷认为，元宇宙不是物理宇宙的替代品，数字艺术也不是物理艺术的替代品，基于数字孪生的新文化在元宇宙中大有可为。产业界应继续开发可编程的、具有拓展性的艺术空间，从而使元宇宙能够赋能人的创造力，而不是减少创造力。元宇宙使得艺术家可以与全世界的同行们交流，不受地理的限制。而由于区块链技术让每一笔交易都可见，所以信用在元宇宙非常重要。

元宇宙概念的提出,使得人们第一次整体性看待过去二十年分别发展的互联网、人工智能、区块链和安全计算等技术及其形成的各类虚拟系统,从而揭示出人类政治、经济和生活已经被各类虚拟系统重新组织,形成了大量基于数字化的、复杂的新关系,颠覆了原有现实世界的权力、权利和关系。与会专家一致认为,赛博空间已经与现实世界的既有秩序、组织和结构形成了张力、也对现实世界的治理提供了全新维度的挑战。这些张力和挑战已经引起了人性挑战、社会撕裂、政治动荡、金融风险等一系列问题,并仍在酝酿新的隐患。因此有必要重新审视目前分散但有联系的各个虚拟空间对现实世界的冲击。虚拟空间中缺乏制度共识和构建、虚拟空间和现实世界制度的冲突,已经成为下一代互联网迫切需要解决的议题。

## （五）专题论坛报告

### 1. 专题论坛一：在相互依存的全球数字时代弥合数字鸿沟

专题论坛一以“在相互依存的全球数字时代弥合数字鸿沟”为主题,围绕南南合作将如何帮助确保不会有任何一个国家掉队并帮助较贫穷国家提高其人工智能技术治理能力,如何使全球人工智能治理和数字合作机制更具包容性,如何通过制定标准、知识交流、能力建设和加强数字合作来确保南半球国家获得公平待遇等问题展开讨论。

巴西总领事馆副总领事兼科学、技术和创新主管,联合国大会和平与安全、人道主义和法律事务高级顾问和团队负责人尤金尼奥·巴尔加斯·加西亚认为不应该赋予人工智能决定生死的权利。人工智能体系的开发者可能只从技术上进行了考虑,但没有社会文化问题方面的深刻认识,因此不应在不考虑社会大背景的情况下就随意应用人工智能。

LIRNEasia 智库主席(斯里兰卡)罗汉·萨马拉吉瓦认为对话很重要,甚至在某种程度上可以影响实际法规制定。通常情况下,政府会做出最终决策,而社会和私营部门在内的所有其他人很难参与到其中。由于知识水平的差异,这可能会导致决策失败。为了解决这一问题,需要更加广泛的磋商。

英国皇家国际事务研究所亚太项目副研究员瓦苏基·沙斯特里指出可解释性和有条件性至关重要。可解释性指的是政府应向其民众解释人工智能政策的作用范围、目标、和影响的同时还应说明数据是如何传播和收集的。有条件性是指应杜绝人工智能的恶意使用,避免其对人类造成伤害。

卡内基国际和平研究院技术与国际事务项目研究员南吉拉·桑布利认为联合国

是唯一一个可以实现人工智能跨代际参与的国际体制。为此,联合国要做好准备来引导这些对话。

瑞典驻华使馆科学创新处参赞南南·伦丁表示人工智能治理是资源密集型的。需要投入必要的精力和资金,同时,也需要互谅互信。只有在弄清如何有效进行人工智能治理多方对话之后,才有机会产出有意义的成果。

与会专家认为人工智能治理已成为一项重要的全球议题,各国应予以重视,并且应在联合国框架下积极推动人工智能治理领域的国际对话与合作。

### 2. 专题论坛二：人工智能伦理与治理中的跨文化合作

专题论坛二以“人工智能伦理与治理中的跨文化合作”这一主题,围绕如何在人工智能治理与伦理的合作基础上实现可持续的人工智能,文化在这种合作中发挥了怎样的作用,如何改善不同文化之间的不信任和误解等问题展开讨论。

剑桥大学生存风险研究中心联合主任、未来智能研究中心项目主任肖恩·欧·海格缇指出人工智能已经对人类生活的世界造成了很大的影响,未来也仍会如此。人工智能给人类带来的许多挑战,不会因文化或国界而异。所有的文化和国家都面临着这些挑战,而解决的方法就是要合作。来自不同背景的人们有着不同的需求和优先要务,所以合作并非易事。只有通过交流意见、相互理解、互换立场,跨文化合作才能水到渠成。

联合国秘书长数字合作高级别小组秘书处前执行主任阿曼迪普·辛格·吉尔认为人工智能合作应建立一个跨学科的、包容的集体。一是当下存在着普世性人工智能与具体文化背景人工智能的矛盾。第一种方向契合了人类的共性,比如人工智能治理中有联合国《世界人权宣言》和其他治理相关的文件等。但人工智能也有着具体的使用情境,例如对于数据收集和治理,各国有自己的框架、标准与治理方式。需要调和这样全球性与本土化的矛盾。二是跨文化合作是一个难点。如果没有法律强制性的规则实施,推动出台人工智能治理条约的进程将会较为困难。所以必须要确保共识性的规则能够适应每一个文化语境,因此需要跨文化的对话来进行指导。三是人工智能在模仿和理解人类智能上是有文化区别的。要保持主观能动性与文化敏感性,以文化敏感的方式去理解人工智能。

埃因霍芬理工大学教授、英国艾伦·图灵研究所研究员文森特·穆勒表示在人工智能伦理和跨文化合作中需要结合情况分析,哪些情况下需要宽容,哪些情况下则需要进行对话和批评。有时划分和平衡这两种情况并不容易,可能会走向批评过

多或宽容过多的极端。尊重合作方文化的关键就在于把握好批评对话和宽容之间的平衡。

联合国教科文组织人工智能伦理特设专家组主席艾玛·鲁特坎普·布洛姆表示通过文化合作来推进全球人工智能伦理建设非常必要，这不是通过一刀切或从上而下的方式，而是从下而上的携手合作。人工智能技术影响着全人类，推动其可持续发展，就必须推动国际合作。此外，对于数据的所有权，必须要平衡以权利为本和以义务为本的价值观，并根据不同的监管环境和价值观来调整做法。坚持使用和支持国际公认的机制，并且始终尊重以不同背景下国际公认价值观为支撑的国际商定机制，坚持所有的文化都应该积极参与到人工智能治理机制和生命周期当中，把跨文化合作自下而上的带入到治理框架当中。

新加坡管理大学法学教授、人工智能与数据治理中心主任马克·芬德利认为人工智能治理中对安全和可靠性的考量较多，而对伦理方面的考量较少，人工智能伦理框架也往往不会被视作一个监管性的框架。从文化角度来看，要让伦理对话被更多人听到，既要有文化背景下适当的机制和框架，也需要普世性、全球性的概念和监管框架。对此，首先要建立人与人之间的信赖，人和机构（机制）之间的信任。其次在制定人工智能伦理原则方面的合作须要将人工智能在各个领域方面的应用考虑在内。不单在跨文化的角度上探讨，也要在全球社群的角度去探讨。

联合国秘书长办公室前 AI 主任、剑桥大学未来智能研究中心外籍研究员丹妮特·加尔认为人工智能的伦理和治理核心就是跨文化合作。当下文化的国际化程度在上升，跨文化合作增加。但忽略了实现真正跨文化合作的两个方面：交流意愿和相互尊重。首先，应进行关于文化概念的全球合作。其次，通过跨文化合作了解不同文化对人工智能的使用和监管，使其跨越地缘政治边界。最后，应在相互尊重、建立信任上投入更多的时间和精力，从而推动跨文化合作。

与会专家一致认为人类本质上是互相依存、互相关联的，面对人工智能带来的全球性挑战时更是如此。对于生活在不同文化背景下的人们而言，文化差异是客观存在的，但这些文化差异更多地体现为表层差异而非根本性分歧。因此，尽管全世界不同地区对人工智能伦理的理解和认知有所不同，但基于人类的共识，人们仍然能够跨越文化差异认可根本性的人工智能伦理原则，建立起共识性的国际价值观。要进一步促进人工智能的可持续发展，把握好普世性与差异化的文化伦理平衡点，推动以相互尊重、加强信任、包容对话为基础的人工智能治理跨文化国际合作势在

必行。

### 3. 专题论坛三：人工智能与气候行动

专题论坛三以“人工智能与气候行动”为主题，围绕人工智能如何帮助应对气候变化，如何权衡人工智能与气候变化并推动两者的创新与合作，如何推动应对气候变化的跨学科解决方案等问题展开讨论。

清华大学能源环境经济研究所所长张希良指出，未来，能源管理包括国家新能源政策和气候政策制定等方面，将出现较大变化。目前基本依靠模型做决策，未来可能向数据导向型决策转变。这种转变不仅体现在电力系统方面，还体现在未来的管理和政策决策方面。在联合国气候治理问题上，科学是非常关键的。联合国在为各方达成协议、取得共识等方面起到了非常大的作用。气候变化问题本身有很大的不确定性，各成员国之间利益也非常不一致，同时还需要考虑效率问题，低成本应对气候变化。而且，各国还有一个共同的责任，就是减排。在此过程中，应注意保密性问题，因为建设碳市场，将来全球的碳市场数据可能要公开透明，但是有一些数据涉及到商业秘密，甚至国家主权，怎样才能既实现数据可靠，又充分保护企业商业秘密，维护国家主权？在这方面，人工智能、计算机科学能发挥很大作用。

复旦大学大数据研究院副院长、能源经济与战略研究中心主任吴力波认为人工智能的应用需求需要计算能力的支撑，而从计算资源而言地区差异性巨大。拿中国来举例，沿海地区人工智能产业集聚、需求最强烈，因为很多电商公司都坐落在那里。但是大部分的可再生能源设施都建设在中国西北部。随着东南沿海对人工智能需求的增大，需要考虑如何优化东西部的算力资源，同时使东部的电力资源请结合，从而使该地区能源供应更有利于应对气候变化。

墨卡托全球公域和气候变化研究所教授费利克斯·克鲁齐格认为义务数据共享对创造良好的政策环境有着巨大的帮助。一些城市有时仍然无法访问诸如航空及交通路线图这样的数据，即使这些数据就在数据库里而且对社会发展非常有益。

BCG Gamma 董事总经理、全球资深合伙人西尔文·杜兰顿认为在 COP26 之后，对公司、政府以及个人而言，最大的问题就是如何减少碳排放。此外人工智能很快就会被广泛应用，同时对环境产生巨大影响。这主要涉及两个问题，一是正在发展的人工智能未来是否会产生大量碳排放；二是人工智能对地球的影响究竟是好是坏。

商汤智能产业研究院院长田丰指出为助力产业元宇宙的可持续发展，人工智能



与数字经济应持续降低能耗，提高能效，推动数字产业化的绿色转型，即“绿色人工智能”、“绿色元宇宙”。人工智能伦理必须以人为本，也必须以环境为本，其底线应包括推动产业持续性发展、科技可控发展。

与会专家一致认为 21 世纪以人工智能为代表的前沿科技将成为实现可持续发展目标的有效工具与创新路径。人工智能将会在帮助全球应对气候变化等共同挑战上发挥积极作用，成为强有力的解决方案的一部分，但前提是每一个人都可以平等的被人工智能所惠及，而所有的国家都对人工智能的开发和使用享有话语权。

#### 4. 专题论坛四：人工智能与国际安全

专题论坛四围绕“人工智能与国际安全”，从理解人工智能对全球战略稳定性的影响和探索降低人工智能相关安全风险的途径两个方面展开讨论。

第一环节主题聚焦理解人工智能对全球战略稳定性的影响，清华大学战略与安全研究中心秘书长、社科学院国际关系学系教授陈琪主持，中国军控与裁军协会副会长兼秘书长李驰江，清华大学战略与安全研究中心客座研究员李睿深，俄罗斯政策研究中心、新兴技术和全球安全项目主任瓦季姆·科久林，日本拓殖大学国际学部教授、海外事务研究所教授兼副所长佐藤丙午，日内瓦安全政策中心全球和新兴风险部主管让-马可·瑞克利等嘉宾出席论坛，就大国关系不断变化的背景下，人工智能如何影响全球主要国家战略稳定性及地区内其他国家提出看法。

第二节聚焦探索缓解人工智能相关安全风险的方法，人道主义对话中心资深顾问埃里克·里查德森，未来战争分析师、印度陆军信号兵团第 57 任中校（退役）R.S. 潘瓦尔，科克大学计算机科学与信息技术学院教授巴里·奥沙利文等嘉宾出席论坛。专家认为需要建立信任措施来预防和控制人工智能开发和部署带来的风险，以及将人工智能合理且正面地应用在国际安全领域。

与会专家一致认为，人工智能的应用将会给全球战略带来极大的不确定性，国际社会应致力于提升人工智能的治理和道德规范，积极探索降低和预防人工智能安全风险的合作路径。

#### 5. 专题论坛五：人工智能与算法公平

专题论坛五围绕“人工智能与算法公平”这一主题，聚焦计算机科学家与社会科学家对人工智能产品的设计、开发以及应用方面存在的不同看法，如何推动不同领域的科学家展开跨学科研究，以及如何消除人工智能中存在的偏见问题和透明度的提升问题。

清华大学产业发展与环境治理研究中心主任陈玲在发言中指出人工智能的合作与治理是一个全球性的问题。目前普遍认为算法只是技术和工具，而如何使用这个技术和工具，人类需要达成共同的规则和共识。不可否认的是，人工智能算法在现实中，在很多领域里都引起了公平性争议。然而，提高算法公平必然就会损失一部分效率，因此算法公平和效率其实是一个权衡问题。针对算法歧视这一问题，来源有三方面：数据，算法和人类行为，本质上是技术秩序的集中体现，无法根除。但是政府对算法歧视进行干预和救济是必不可少的。而对于算法公平这一问题，需要注意的是，不同的公平观是内在冲突的，不可兼得的。既想要起点公平、过程公平，又想要结果公平，这是不可能实现的。在相互冲突的观念和计算方法里如何进行抉择？答案其实就是找到可接受公平。在算法公平里寻求一种最低限度的、共识性的可接受公平，就是进行全球算法治理的起点。

亨氏信息系统与公共政策学院机器学习系数据科学与公共政策中心主任伊德·加尼认为首先需要定义隐私问题。隐私和信赖缺失问题取决于数据用途，而非算法需要数据，算法本身只是手段、而非目的。如果问题是实现什么样的价值观，那就要提高关于算法背后价值观的透明度。如果在价值观上未达成共识，仅改代码是没有用的。

科幻作家陈楸帆指出从历史与哲学的角度讲，人类现在走的是一条主客体二分的道路，期待通过人工智能视角，跨越二分。这种二分不仅仅是存在于人类的性别、种族、阶层、国家、文化之间，也包括在人类与其他物种之间如何和谐共生的一种行星级智慧，期待能够以此解决当下以及未来所面对的所有文明层面的重大挑战与终极考验。

麻省理工学院斯隆管理学院运营管理助理教授丹尼尔·弗洛伊德指出公平性问题是可解决的，比如如何进行资源公平分配。如果不考虑公平元素，技术能解决的问题是 0%，考虑公平的情况下，解决问题的占比是 100%。

联合国妇女署项目和伙伴关系专家毕文韬指出需要构建包容、变革性的、负责任创新生态体系。应运而生的人工智能就是这一生态体系的产品。如果这一体系能够更加包容、负责，那么就可以拥有更好的人工智能产品。另外，在行业内需要为女性赋能，促进人工智能朝更加包容多元的方向发展。

与会专家一致认为应推动计算机科学家、社会科学家之间的多学科研究，促进算法公平，进一步消除人工智能中的偏见问题。



## 6. 专题论坛六：国际数字治理与企业可持续发展

专题论坛六以“国际数字治理与企业可持续发展”为主题，聚焦在全球数字治理形势下企业面临的风险与挑战。

美团副总裁、首席科学家夏华夏指出，数据是数字经济的“石油”和数字化社会的“血液”。数据治理的方向不应该是限制数据流动，而是在合规的治理下，保障数据有序地自由的流动，是以促进数字经济发展和个人信息合理利用为目的的。数据治理作为一项系统性工程，需要依托法律、技术、市场等多元化管理手段，也需要国际社会、国家、企业、个人的共同努力，才能保证数字市场生态的有序发展。

华为公司人工智能战略与产业发展副总裁秦尧表示，人类已经迎来第四次工业革命，人工智能的广泛应用使人类加速进入智能世界。目前在实际应用中，安全及可靠性问题对人工智能的应用产生了极大的挑战。社会大众对人工智能应用风险的担忧持续增强使得人工智能治理变得尤为关键。人工智能应该服务于保护环境、提升人类福祉和解决行业问题，建立一个开放包容、全球共治的可信赖人工智能。同时，应打造国际多边治理机制、完善标准和认证能力、构建分层治理框架、促进包容性发展是构建负责任、可信赖的人工智能治理体系的关键。

京东探索研究院院长、澳大利亚科学院院士陶大程指出，人工智能系统的不稳定性容易导致潜在安全问题，而可解释性的缺乏也限制了人工智能更广泛的应用与赋能，未来人工智能系统该如何在使用数据的同时保护用户隐私则成为大众关注的焦点，所以发展可信人工智能刻不容缓。具体来说，应该从稳定性、可解释性、公平性与隐私保护等四个维度来构建可信人工智能，可信人工智能一体化研究会将是未来的重要趋势，可信能力评测也将成为未来人工智能产业落地过程中的重要一环，总体来说从理论与实践层面持续开展可信人工智能研究将推动人工智能产业新浪潮。

英特尔研究院副总裁、英特尔中国研究院院长宋继强认为，人工智能在传承历史文化、保护动物自然和守护人类健康等方面发挥了巨大的作用，是实现联合国可持续发展目标的重要技术手段。同时，人工智能本身的可持续发展也面临着挑战，全球数据量呈指数级增长，大规模人工智能模型消耗的能源是不可持续的。法律法规、国际标准、行业公约和企业自律是构建人工智能治理体系，实现健康与可持续发展的四个关键。企业要做到履责、包容、可持续和赋能，通过技术创新，推动社会变革，促进技术包容性，改善健康和安全的实现碳中和，对社会、企业、地球环

境创造积极的影响。

清华大学法学院院长、智能法治研究院院长申卫星指出数据作为一种资源，是数字经济的“石油”。为更好地发挥数据价值，他建议完善法律法规，在用户和平台之间合理分配产权，可以借鉴传统要素（例如土地）法律制度化的过程，建立数据要素所有权和用益权两权分离的确权模式，即在用户享有数据所有权基础上，平台基于付出的资本、技术和劳动，获得用益权。目前我国数字经济发展需要建立合理、充分的治理体系和框架，实现技术、伦理、法律“三驾马车”驱动。一方面，要“平衡”创新与监管。例如，新出台的《个人信息保护法》强调，不仅要加强个人信息保护，还要加强个人信息的合理利用。另一方面，要注意可持续发展，避免走“先污染后治理”的老路。

达闼创始人、董事长兼 CEO 黄晓庆认为人类对信息权益的关系和控制，最成熟的产业就是通讯行业，第二成熟的行业就是银行业，第三成熟的行业是出版业。对人类来说，对于数据管理和控制，在一些领域已经有建立起来的体制和机制，值得参考。一定要建立起公共的计算体系，在零信任的基础之上，得到所有社会主体的公平有效使用。未来，社会整体的“元宇宙化”是可能遇到的挑战。在法规上，中国应该领先于世界，去建立端到端的人工智能数据安全监管法规，没有监管机制未来会面临巨大挑战。

地平线首席生态官、副总裁徐健认为，数字对于人类生活带来很大机遇和挑战，一个非常重要的领域是汽车人工智能芯片。当前智能汽车自动驾驶发展非常迅猛，数字人工智能对于整个汽车行业带来很大的机遇，也面临着挑战。实际上产业共同治理的格局还未充分形成，在智能汽车领域里不同参与者如何共同的迭代优化，如何对数据进行更好的管理目前还处于摸索阶段。如何定义好数据所有权，定义好数据在人工智能训练、传输、安全中所应遵循的规则，是整个产业界遇到的共同话题。数据应加快流转，一是在智能汽车领域里要关注边缘计算，很多智能驾驶的问题应放在端侧解决；二是要注意考虑硬件加密对于打造自动驾驶安全可靠的作用。

清华大学中国电子数据治理工程研究院院长、互联网治理研究中心主任孟庆国认为，数字经济时代要平衡发展和安全的关系。数据的归集、流动、分享和应用都存在很多风险，应当思考这背后的机制和原则。目前，存在一个政府、企业、行业和社会都发挥作用的综合治理体系。政府既要制定产业规划，鼓励发展，也要立法，制定标准以提高公平性、包容性；企业应该自律，注意合规性建设，保证平台开放

性，实现公平交易和保障安全；行业要提出大家都接受的倡议，开展合规集体行动；社会要发挥监督作用。对于数据跨境流动，跨国企业要做到“双向合规”，要面临所在国、目的国和自身商业数据保护的“三重困境”，而全球分散化、圈群化的跨境数据规则体系也愈加使得企业无所适从。面对挑战，中国应该更多参与到多边国际组织里，参与跨境数据治理体系和规则的建立。在国内数据立法和细则制定中，应引入企业更多的参与，同时采取更积极和包容的策略来面对企业的跨境数据需求问题。

### 7. 专题论坛七：人工智能与社会发展

专题论坛七围绕“人工智能与社会发展”，就如何推动负责任、值得信任的人工智能发展，并以此实现可持续的社会发展等问题展开讨论。

清华大学公共管理学院教授、常务副院长，全球可持续发展研究院执行院长、科技发展与治理研究中心副主任朱旭峰表示人工智能在社会发展中带来的挑战主要表现在三个方面。第一，是技术的成熟度；第二，是所有技术发展都要面临的一个共性通用问题，即对社会影响的两面性和颠覆性；第三，是与人工智能相关的伦理问题。这些方面都需要进一步归类与精细化。考虑到国际政治因素，任何技术的转移要考虑的不仅是技术先进国对技术落后国转移的意愿，技术落后国是否愿意吸收技术先进国的技术也是不确定的。比如大数据分析或金融领域的人工智能技术等，如果国家之间本身不信任，技术落后的国家可能也不会接受这些技术，会担心自身金融系统被发达国家所掌握。

慕尼黑工业大学人工智能伦理研究所主任、清华大学人工智能国际治理研究院学术委员克里斯托夫·吕特格指出在之前的技术哲学中，核技术被普遍认为风险很高。一旦爆发核战，整个人类可能会被彻底毁灭，所以各国对此认识有共识。但对于人工智能技术影响的认识并不是那么清楚。因此人工智能在这方面的影响就像黑匣子一样，尽管存在一些隐忧，但不确定会在未来什么时间点造成巨大的挑战。从各方角度来说，技术转移应该是一个互学互鉴的过程，技术人员与非技术人员之间需要相互学习，不同文化与文明之间需要加强了解和借鉴，需要将不同国家对使用人工智能的顾虑与抵触都纳入考量。因此，有必要开展更多关于伦理和全球治理方面的讨论。

中科院自动化所中英人工智能伦理与治理中心主任、清华大学人工智能国际治理研究院首席专家曾毅认为实现可信赖的人工智能不仅要技术的层面下手，也需

要从社会角度出发。在谈论人们是否信赖人工智能的时候，也需要探讨不同地域和文化之间的信任。由于各个国家的文化价值观不同，一个国家是否信任另外一个国家开发的人工智能产品，这也是讨论信任问题中需要关注的层面。在技术转移过程中文化互信及对不同文化的尊重是至关重要的。如果要进行技术转移，必须要使产品适应新的文化环境，去理解对方的需求和顾虑，这样才能赢得对方的信赖，使其使用被提供的产品或者平台。此外，基础设施的共享某种程度上要比产品共享更加重要，这可以让对方根据自身国情、人力资源与需求去创造所需要的产品和服务。

清华大学产业发展与环境治理研究中心执行理事、原中信证券股份有限公司董事总经理德地立人表示在一些发展中国家，如果新兴技术人工智能等不能得到很好控制的话会很容易扩散，由此出现与周边国家的冲突。控制好新兴技术向发展中国家的输出，对可持续发展是有好处的。在这方面，国际组织需要有一定的制度来进行监控和管理。不然，可能带来的是对整个地区或对全人类的悲剧。应对技术先进国与落后国之间的不平等需要长期的准备与努力，要让不同国家之间能够达成共识，同时最关键的是要在教育领域投入更多精力，让新一代的儿童与年轻人能够接受到平等的人工智能教育。

北京市大数据中心数据管理部负责人贾晓丰人工智能背后的本质、运行机理及其以及带来产业链条的变化。人工智能对于可持续发展，不论是从信息对称、资源要素更精准的匹配，还是作为生产生活与科研军事各方面的新一代生产力，都起到了显著的推动作用。人工智能不是要把以前的事情做得更好，而是要从根本上改变当前的治理体系，从而改变目前的生产生活方式，乃至未来的生存方式。实现责任和可信赖的人工智能发展需要在三件事上凝聚共识。第一，国际社会要有对人工智能基本准则和规律达成有动态共识的机制，包括制度体系、政策、能耗、技术路径等。第二，人工智能从基础理论到工程化，再到产业化的过程并不是单线条发展的。因此要推动跨学科交叉，并且除了内部的孵化，还要有外部的扩散。第三，需要防范未来的系统性风险。不仅要解决今天面临的问题，还要为5-10年后以及更远的未来做进一步的数字增量布局。

慕尼黑工业大学人工智能伦理研究所研究人员亚历山大·克里比茨认为人工智能技术能够显著提升地区生产力，在非洲部分没有实现工业化的地区已经率先的实现了数字化，发展非常迅速。因此，人工智能需要找到任务和愿景，理清借助人工智能要实现可持续发展目标。人工智能在政务中的使用需要更加谨慎。因为运用

人工智能在政务之中涉及到合法性的问题，必须确保政府的行为是建立在民众同意的情况下，而且还必须是符合宪法的。同时，人工智能需要在可解释性方面有近一步提高，否则很难构建公众对人工智能技术的信赖。

与会专家一直认为人工智能技术在赋能千行百业的同时，也带来了安全、隐私、公平等问题和挑战。解决这些问题需要加强全球协作，出台相关政策法规，推动构建平衡包容的人工智能治理体系，促进人工智能可持续、负责任的发展以更好地造福人类社会。

## 五、2022 人工智能合作与治理国际论坛

2022 人工智能合作与治理国际论坛将于 2022 年年底举行，会议信息即将发布，敬请关注。

人工智能合作与治理国际论坛网址：

<https://www.tsinghuaaiforum.org/>

清华大学人工智能国际治理研究院网址：

<https://aiig.tsinghua.edu.cn>

联系：i-aiig@tsinghua.edu.cn



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