



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



国际支持机构
Supporting International
Organization

2020清华大学人工智能合作与治理 国际论坛白皮书

2021年3月

目录

一、论坛介绍	03
二、论坛嘉宾	04
三、论坛成果	06
四、论坛观点总结	09
（一）开幕式讲话	09
（二）主论坛报告	11
（三）分论坛报告	12
五、第二届清华大学人工智能合作与治理国际论坛	21

一、论坛介绍

在此次抗击新冠肺炎疫情中，人工智能（AI）技术在病毒检测和诊断、疫情追踪以及日常监测上发挥了关键支撑作用。然而，伴随人工智能技术的广泛应用，也引发了对于经济社会发展中利益与潜在风险平衡问题的担忧。人工智能技术可能带来对个人隐私和不平等的挑战，并导致新型歧视、偏见、社会分割等诸多问题。

近年来，各国政府、非政府组织、科研团体和私营机构相继发布了人工智能原则文件，如何平衡人工智能开发与防控风险一直是讨论的焦点。随着新冠肺炎疫情在中国逐步得到控制，我们应抓住时机，结合此次抗疫，反思人工智能带来的利处，以及在不同文化背景下如何增进各利益攸关方对人工智能治理的相互理解，推进国际人工智能治理框架的建设进程，助力人工智能构建可持续和包容的未来，避免加剧挑战和产生数字鸿沟。

在此背景下，由清华大学人工智能国际治理研究院主办，联合国开发计划署作为国际支持机构的首届清华大学人工智能合作与治理国际论坛于2020年12月18日至19日在中国北京召开。本次论坛以“后疫情时代的人工智能国际合作与治理”为主题，聚焦人工智能技术带来的机遇和挑战，分享人工智能合作与治理见解，交流国际治理智慧，提出未来人工智能治理研究议题。

二、论坛嘉宾

中方嘉宾（按报告顺序）

王钦敏	第十二届全国政协副主席，全国工商联原主席
邱勇	清华大学校长
杨斌	清华大学副校长、清华大学人工智能国际治理研究院管委会主任
许惊	科技部战略规划司司长
王小龙	外交部国际经济司司长
朱秀梅	工信部科技司副司长
姚期智	图灵奖得主、清华大学交叉信息研究院院长
张钹	中国科学院院士、清华大学人工智能研究院名誉院长
傅莹	清华大学人工智能国际治理研究院名誉院长、前外交部副部长
张亚勤	清华大学智能产业研究院院长
薛澜	清华大学人工智能国际治理研究院院长
王磊	外交部网络事务协调员
梁正	清华大学人工智能国际治理研究院副院长
于洋	清华大学人工智能国际治理研究院国际学术交流项目主任、清华大学交叉信息研究院助理教授
夏华夏	美团副总裁，首席科学家
史军	商汤科技亚太业务事业群总裁、战略规划副总裁
洪小文	微软全球资深副总裁、微软亚太研发集团主席、微软亚洲研究院院长
宋继强	英特尔中国研究院院长
唐剑	滴滴智能控制首席科学家
王晶晶	公众环境研究中心副主任
孔秋实	杭州市萧山区数据资源管理局副局长
张敏	清华大学计算机系人工智能实验室副主任
蒋燕	旷视副总裁
曾玉	和玉资本管理合伙人
曾毅	中科院自动化所中英人工智能伦理与治理中心主任、北京智源人工智能研究院人工智能伦理与可持续发展中心主任
朱旭峰	清华大学公共管理学院副院长、清华大学全球可持续发展研究院执行院长
郑治泰	地平线首席战略官
肖茜	清华大学人工智能国际治理研究院副院长
李彬	清华大学社会科学学院教授、清华大学战略与安全研究中心客座研究员
郝英好	中国电科发展战略研究中心高级工程师
李驰江	中国军控与裁军协会副会长兼秘书长
黄伟豪	香港中文大学数据科学与政策研究所所长
邓亚峰	奇虎 360 人工智能研究院院长
曹建峰	腾讯研究院高级研究员
吴沈括	北京师范大学教授

外方嘉宾（按报告顺序）

白雅婷	联合国开发计划署驻华代表
法布里齐奥·霍奇尔德	联合国副秘书长、联合国秘书长特别顾问
赵厚麟	国际电信联盟秘书长
约兰达·吉尔	国际人工智能协会前主席
李在摄	国际电信联盟标准局局长
凯·菲尔斯-巴特菲尔德	世界经济论坛人工智能和机器学习项目负责人
埃菲·瓦耶娜	世界卫生组织健康领域 AI 伦理与治理专家组组长、苏黎世联邦理工学院大数据与人工智能伦理学教授
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
阿德里安·韦勒	剑桥大学工程系高级研究员、图灵研究所人工智能项目主任
罗欣顿·麦德拉	加拿大国际治理创新中心总裁
托马斯·戴文	联合国儿童基金会创新主任
马克斯·泰格马克	麻省理工学院物理学教授
张薇	联合国开发计划署助理驻华代表
克里斯蒂安·古特曼	北欧人工智能研究所执行董事
罗罕·萨马拉吉瓦	斯里兰卡技术基础设施智库主席
米拉·罗曼诺夫	联合国全球脉动行动数据隐私专家
南南·伦丁	瑞典大使馆创新参赞
约翰·艾伦	布鲁金斯学会主席
尼尔斯·吉尔曼	博古睿研究院项目副总裁
克里斯·梅塞罗	布鲁金斯学会研究员
汤姆·斯蒂芬尼克	布鲁金斯学会访问研究员
戴文德	联合国开发计划署驻华副代表
齐亚·汗	洛克菲勒基金会高级副总裁
杨子健	新加坡信息通信媒体发展管理局助理首席执行官
文杰	韩国延世大学未来政府研究所所长
鲁德拉得·米特拉	印度 Omdena 创始人
苏里纳·舒克里	马来西亚数字经济发展局首席执行官
赛勒斯·霍德斯	未来社会 AI 倡议联合创始人兼主任
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
沃伦·海罗	南非 Webber Wentzel 律师事务所首席信息官
丹尼特·加尔	剑桥大学智能未来研究中心研究员
文森特·穆勒	荷兰埃因霍温科技大学教授
罗曼·扬波夫斯基	美国路易斯维尔大学教授

三、论坛成果

（一）形成人工智能国际治理的专家共识

第一是通过技术创新，实现负责任的人工智能技术，包括数据分析模型，通过预测来防范疫情，以及提升个人健康水平的应用开发等。在这一过程中，除了要破解数据质量和算法算力难题之外，还要关注应用中所伴生的一些亟待解决的突出问题，例如对公共安全、伦理道德、社会就业、市场竞争和隐私保护等方面所带来的影响和冲击，以及在数字政府、智慧城市等建设过程中所出现的治理问题。

第二是加强国际合作，在国际层面上进行充分的对话与交流，因为没有任何一个单一的国家或公司能够设计出全面、符合所有人预期的指导方针来管理人工智能的发展以及它在全球引起的连锁反应，必须为人工智能创建一个可行的国际合作和治理框架，动员各方积极参与和努力推动。特别是从此次疫情应对中发现，国际合作和多边主义至关重要，当前必须从全球层面来探讨人工智能相关的话题，为此可以在人工智能领域建设包容性强的国际治理委员会，共同去研究、探讨、吸纳各方面好的建议和意见，形成共同的国际规范。

第三是在推动人工智能大规模使用同时，要制定合适的治理准则以确保相关技术的安全性问题。目前在人工智能推动社会进步同时，存在滥用数据和侵犯个人隐私等风险。必须谨慎对待，确保人工智能可以推动人类发展，避免加剧不平等或产生新的挑战，从而阻碍全球可持续发展目标的实现。为了应对这样一大挑战是，即人工智能技术发展过快，而治理准则发展是缓慢的、渐进的情况，需要进行“敏捷治理”。人工智能可以学习已有的国际机制，未来不断加强人工智能领域的国际合作，可以在合作过程中找到各方的共同点，并不断探索、沟通，一同寻找人工智能共同的准则和规则，引导人工智能的健康发展，为人类造福。

（二）总结人工智能在新冠疫情防控上的先进经验

论坛聚焦于人工智能技术在疫情期间精准识别、精准防控、精准施策、复工复产等方面发挥的重要作用，特别是各国最佳实践，以及在这一过程中出现的问题和解决思路。

然而由于各个国家的保障体系和数字化程度不一，数字转型和治理是非常紧迫

的任务，有必要让各国进行清晰的合作和协调。在人工智能技术的发展过程中，也要注意伦理方面的问题，人工智能伦理风险的边界和核心要求是发展负责任的人工智能，通过技术的提升，以及政府法律有机结合来规范技术的使用。

此外，人工智能技术不应仅仅追求即刻的优化，同时也要思考系统怎样能够更加强健。在疫情下卫生健康方面，我们应开始关注和思考怎样实现个人和集体福祉间的平衡，如应该如何分享健康数据，从而使得每个人都从中受益。

（三）关注人工智能对于促进可持续发展的作用

目前人工智能治理在促进可持续发展方面并不平衡，主要集中在教育和卫生领域，而在气候、零饥饿、生物等很多方面缺乏研究。专家们提出 17 个联合国可持续发展目标（SDGs）并非相互独立的目标清单，AI 技术在对某一 SDG 目标产生影响时，也势必会影响到其他目标。因此应将可持续目标视为一个整合性的框架，在 AI 技术的推动下实现变革性的长久转变，并由各国开放探讨 AI 策略，实现在不同文化背景下的共同诉求。

现在各国间仍然存在数据鸿沟问题。用于训练 AI 的数据生产并不平等，主要来自富裕国家和人群，机器对于低收入人群的了解仍然不足。因此，应当合理关注中低收入情境。建立有效的政府与企业的制衡体系，考虑引入第三方进行独立评估。关注权益与偏见问题，对不可接受的 AI 使用划定红线。并且应当设置确保社会法律问责与技术问责能够实施落实的工具，包括审核、记录机制等。最后，政府在将人工智能视为赋能工具的同时，应划定清晰的发展路径和实现框架。

（四）探讨人工智能对国际安全的挑战与机遇

人工智能对国际安全存在潜在挑战主要表现为，世界各国围绕人工智能进行不可控的军备竞赛和人工智能武器的滥用。人工智能技术的滥用或者偏差有可能冲击战略稳定、改变交战规则、抬高军事风险。在国际安全领域，人工智能技术的训练数据安全的应对，至少应遵循秉持多边主义、兼顾安全发展和坚守公平正义三项原则。在人工智能赋能平台上，各国应当在多方共同参与的基础上，在联合国框架内，按照比例原则和区分原则，通过人类共同监督来避免意外的冲突升级及其带来的智能武器滥用。

中美两国在建立人工智能安全国际规范方面应发挥积极的作用。在政府层面对

话和非官方论坛中进一步探讨人工智能与国际安全的内在关系，特别是针对人工智能所构成的国家安全风险，以降低这些风险发生的可能性。抓住当下的机遇，围绕降低人工智能可能构成的国家安全风险的实际步骤达成共识，规范并引领人工智能技术的发展和运用。

（五）提出人工智能国际治理的合作倡议

当前人工智能伦理准则中最缺乏的是来自社会特别是底层群众和边缘群体的声音，他们对人工智能的担忧和抵触应当得到政策制定者的充分考量，有机会决定想要构建的世界。与会专家认为应当通过“知识民主化”赋权公众，让公众有机会参与人工智能治理层面的解决方案，制定出具有国家特色的人工智能发展模式。

面对人工智能对国际治理带来的挑战，迫切需要大量了解人工智能治理领域的审计人才，以确保治理框架行之有效，同时企业也需要从根本上思考自己的商业模式，才能在这些变化中生存下来。

从全球的视角来看，迫切需要倡导构建一个由全球科学家和相关专家组成的资源网络，支持各国自身的人工智能治理体系的建设。对此不能只关注每个国家自己的数字主权和经济利益，而要更全面地思考负责任的人工智能准入问题，国际间通力合作，才能让人工智能最大化地促进全球社会效益。

（六）阐释人工智能与数据安全的重要内涵

数据和人工智能安全中存在内部安全（Safety）和外部安全（Security）两种。其中，内部安全更偏向内部，针对的是一些无意伤害的预防措施，从系统中保护环境。外部安全更倾向于对外部攻击的防御，聚焦在那些有意对个人、组织和财产施加侵害的行为，从环境中保护系统。数据和人工智能安全与保障面临短期挑战和长期不确定性。短期挑战是人工智能安全相关的内部风险与外部防御问题，与大量隐私数据关联，这些数据往往需要被送到云端进行分析和计算，造成了比原来更严重的数据泄露和安全风险，数据违法行为的跨国行变；长期挑战在于通用人工智能和超级智能具有极大不确定性，从超级智能到技术奇点，再从技术奇点到人类社会的整体生存风险。

世界各国应加强交流、推动合作，以减少脆弱性，在全球范围内倡导并激励数据安全治理工作，将全球范围内数据安全战略联系在一起以造福全人类。世界各国

和组织应联合起来制定更加明晰有效的法律法规，构建统一的人工智能模型与系统的安全评估准则和标准，在全球范围内用更加开放协同的方式进行科学研究和技术交流，推动人工智能技术在安全的基础上更快速的落地，构建更安全、更美好的世界。在企业、国家乃至国际层面建立负责任的人工智能开发和数据使用的道德准则。人工智能与数据产品开发需要遵循“通过设计的伦理”（ethics by design）理念，将公平、安全等伦理价值融入产品设计全周期，并有防止滥用的保障措施。

四、论坛观点总结

（一）开幕式讲话

第十二届全国政协副主席、全国工商联原主席王钦敏致开幕词，并表示人工智能技术的开发利用中，除了要破解数据质量和算法算力难题之外，还要关注应用中所伴生的一些亟待解决的突出问题，例如对公共安全、伦理道德、社会就业、市场竞争和隐私保护等方面所带来的影响和冲击，以及在数字政府、智慧城市等建设过程中所出现的治理问题。作为世界知名的高水平研究型大学，清华大学要发挥在人工智能治理领域的全球学术与政策影响力，以探索学术前沿、服务国家重大需求为己任，不断提升创新能力，努力开拓学术前沿，积极为国家发展建言献策、为人类文明贡献智慧。

清华大学校长邱勇表示，作为人工智能思想的起源之地、社会进步发展的创新之地、人类未来希望的孕育之地，大学要在人工智能合作与治理方面开创新局面、展现新作为：要引领人工智能的发展方向；要不断创新人工智能的理论与技术；要积极推进人工智能全球合作。今年6月，清华大学成立人工智能国际治理研究院，面向人工智能国际治理重大理论问题及政策需求展开研究，努力构建公正合理的人工智能国际治理体系，为应对全球重大挑战提供智力支持。大学作为人类文明的灯塔，代表着人类的美好和光明。越是在不确定的迷雾中，大学越是要对人类掌握自身发明创造的能力和智慧充满信心，对人类发展出最终造福人类的人工智能技术充

满信心，让永恒的文明之火，照亮人工智能发展之路。

联合国开发计划署驻华代表白雅婷致欢迎词。她表示，在此次疫情的应对中，国际合作和多边主义至关重要，必须从全球层面来探讨人工智能相关的话题。在数字时代，数据以指数级速度激增，推动社会进步的同时，也存在滥用数据和侵犯个人隐私等风险。我们必须谨慎，确保人工智能可以推动人类发展，避免加剧不平等或产生新的挑战，从而阻碍全球可持续发展目标的实现。她呼吁要在数字基础设施和教育领域进行投资，确保人们可以从技术创新中受益，增强能力以便于更好地适应由人工智能和自动化带来的新世界。

联合国副秘书长、联合国秘书长特别顾问法布里奇奥·霍奇尔德认为，新冠肺炎疫情加速并且深化了人工智能的应用，使其在健康领域的运用显著增加。尽管人工智能潜力巨大，但也带来了挑战，如我们越来越依赖人工智能的预测能力，这使得我们容易受到系统性错误、网络攻击、致命性自动化武器以及不断扩大的数字鸿沟的影响。他敦促开展国际合作，确保以透明、值得信任的方式使用人工智能，以维护人权和人类尊严，提升我们的安全保障，并促进普惠包容性和平。

国际电信联盟秘书长赵厚麟在视频致辞中表示，对于人工智能市场规则的设立和社会伦理方面的规范，国际治理环境和框架设立非常重要，需要动员各方积极参与和努力推动。他说，国际电信联盟愿与清华大学共同推动全球人工智能技术创新与协同治理，为人工智能惠及全人类作出更大的贡献。

科技部战略规划司司长许倬发表主旨演讲，他指出，中国政府高度重视人工智能治理问题。人工智能治理要以科学技术进步为引领，以增进人类福祉为目标，以确保安全可控为底线，以全球开放共治为理念。期待清华大学能够依托人工智能国际治理研究院，开展深入研究，推动交流对话，集聚全球智慧，形成合作共识。

外交部国际经济司司长王小龙发表主旨演讲，并表示实现对人工智能有效治理，离不开政府、科研、产业、智库等各方面共同参与，协同努力。清华大学人工智能国际治理研究院自成立以来，搭建合作桥梁，开展国际合作，为加强中方同各国在人工智能领域的合作发挥了重要的作用。未来，人工智能治理应把握向好向善发展方向，营造开放合作的有益环境，坚持包容公平的规则制定，秉持共商共建共享的原则，避免“脱钩”和“封锁”，尊重各国的治理主权和法律法规，为人工智能的更新迭代提供空间。

工信部科技司副司长朱秀梅发表主旨演讲，强调要高度重视人工智能的治理，

加强人工智能领域的国际合作，充分凝聚国际社会的共同智慧，实现发展与治理的正向效应。在人工智能治理问题上实现最大公约数，形成共谋发展、共商治理的良好局面。此次举办的清华大学人工智能治理国际论坛，就人工智能国际治理开展深度的对话和交流，具有积极的现实意义，期待通过此次论坛为推动人工智能国际治理体系的建设贡献中国智慧，为国际科技交流与合作增添更多的务实成果。

（二）主论坛报告

图灵奖得主、清华大学交叉信息研究院院长姚期智在主论坛发言中以多方可信安全计算技术(MPC)为例，指出加密技术在数据治理上能够发挥更加重要的作用，并建议将一系列可信算法和相关的真实机构结合起来，通过合作建立可信数据治理的基础。

中国科学院院士、清华大学人工智能研究院名誉院长张钹从当前人工智能的脆弱性谈起，介绍了清华大学人工智能研究院在第三代人工智能方面的最新工作，指出人工智能发展与治理应当相辅相成，共同前进，并呼吁各国团结起来共同发展安全、可靠、可信、可拓展的第三代人工智能技术。

国际人工智能协会前主席约兰达·吉尔认为现今人工智能技术的多样化、深度化发展，推进了科学研究，为解决更多的社会问题创造了条件。她指出要加强人工智能领域的国际合作和数据共享，进一步助力科研发展，尤其是跨学科研究。

国际电信联盟标准局局长李在摄在视频报告中介绍了国际电信联盟在人工智能领域的工作，他指出，各利益攸关方应加强沟通与合作，探讨人工智能如何更好地服务于人类社会，促进联合国可持续发展目标的实现。

清华大学人工智能国际治理研究院名誉院长、前外交部副部长傅莹在主论坛讲话中呼吁在人工智能领域建设包容性强的国际治理委员会，共同去研究、探讨、吸纳各方面好的建议和意见，形成共同的国际规范。她强调，“中国国内在对人工智能技术应用的治理和立法方面已经在采取实际行动。当然，从更大的角度看，这是人类共同的难题，不是一个国家、两个国家能解决的。中美两国乃至国际机构和各国围绕这个话题展开合作至关重要，我们希望也相信这不是一场零和游戏，人工智能最终应使得全人类受益。”

清华大学人工智能产业研究院院长张亚勤在发言中强调了人工智能治理的三大维度，一是通过技术创新，实现负责任的人工智能技术，包括数据分析模型，通过

预测来防范疫情，以及提升个人健康水平的应用开发等；二是加强国际合作，在国际层面上进行充分的对话与交流，以应对最为紧迫的问题；三是推动人工智能的大规模使用，以及同时确保相关技术的安全性问题。

清华大学人工智能国际治理研究院院长薛澜教授认为，目前人工智能发展面临的巨大挑战是——人工智能技术发展太快，而治理准则作为一种社会体系，其发展是缓慢的、渐进的，因此我们需要“敏捷治理”。他指出，“人工智能可以学习已有的国际机制，未来不断加强人工智能领域的国际合作，我们可以在合作过程中找到各方的共同点，并不断探讨、沟通，一同寻找人工智能共同的准则和规则，引导人工智能的健康发展，为人类造福。”

外交部网络事务协调员王磊指出数据安全对于人工智能治理至关重要，中国政府提出的《全球数据安全倡议》是基于联合国等多边平台多年来的讨论情况以及各国数据安全治理实践，就各方关注的重大数据安全问题提出的建设性解决思路。各国应坚持多边主义，共同推进数据安全、人工智能治理进程。

世界经济论坛人工智能和机器学习项目负责人凯·菲尔斯-巴特菲尔德认为人工智能应为人服务，让人们受益于人工智能，并且保持人类对于人工智能的主导性，因此，探讨人工智能伦理至关重要，各方应加强国际合作，制定人工智能伦理准则。

（三）分论坛报告

1. 分论坛一：人工智能在抗击新冠疫情中的作用：经验与教训

2020年12月18日下午，首届清华大学人工智能合作与治理国际论坛分论坛暨高峰论坛——“人工智能在抗击新冠疫情中的作用：经验与教训”成功举办。论坛由清华大学人工智能国际治理研究院承办，清华大学人工智能研究院，清华大学中国科技政策研究中心，清华大学科技发展与治理研究中心，清华大学产业发展与环境治理研究中心共同协办。论坛由清华大学人工智能国际治理研究院副院长梁正主持，研究院院长薛澜教授致欢迎辞。

清华大学人工智能国际治理研究院国际学术交流项目主任、清华大学交叉信息研究院助理教授于洋认为在中国的疫情治理实践中，以人工智能为代表的信息和数据技术参与疫情治理具有“全”和“快”两大特点。中国的疫情治理是由一个AI适应性的政府，以及主动而有能力的企业所形成的一体化的治理共同体完成的。人

工智能技术之所以能够全面快速渗透性的参与中国的疫情治理，主要是由于企业主动参与治理，算法思维的政府，以及在政府和企业中都大量存在的枢纽性部门三个关键因素。

世界卫生组织健康领域AI伦理与治理专家组组长、苏黎世联邦理工学院大数据与人工智能伦理学教授埃菲·瓦耶娜认为，由于各个国家的保障体系和数字化程度不一，数字转型和治理是非常紧迫的任务，有必要让各个国家进行清晰的合作和协调。但是，在人工智能技术的发展过程中，也要注意伦理方面的问题。比如，欧洲政府希望保护民众免受疫情危机的影响，但同时也不违反法治、民主和基本的权利。

美团副总裁、首席科学家夏华夏认为，红外体温识别，无人配送机器人、在线办公、在线教育、直播零售等很多人工智能技术在抗疫中发挥了重要作用，人工智能带来的伦理问题可以用技术的提升去解决，比如可以通过人工智能技术手段让很多数据在端上处理，比较好的保证用户的隐私。还有一些AI的问题，是技术不太容易去解决的，需要跟政府、法规有比较好的结合，通过法律来规范技术的应用。

商汤科技亚太业务事业群总裁、战略规划副总裁史军认为，后疫情时代，人工智能继续成为全球竞争的新焦点。可以看到，各个国家在人工智能技术发展和数据治理等领域发展很快，因此需要进一步完善监管和治理规则。他指出，AI伦理风险的边界和核心要求是发展负责任的人工智能。在治理和合规上企业要先想一步，先做一步。人工智能企业主动监管，建立起动态机制进行管理。

高峰论坛的第一场专题对话围绕“人工智能在抗疫中的国际经验”展开，分享了高科技公司、社会组织 and 地方政府利用人工智能助力抗疫的实践经验。对话环节由清华大学人工智能国际治理研究院副院长梁正主持。

微软亚洲研究院院长洪小文用3个“R”对疫情进行了总结，即“response 回应”、“recovery 恢复”和“reimagining 重塑”。同时指出了人工智能技术在三个阶段所产生的重要影响，如“response”阶段通过线上交流进行远程医疗、办公，“recovery”阶段使用健康码实现快速追踪和疫苗的研发以及“reimagining”即后疫情时代的线上线下混合办公形式的出现等。

英特尔中国研究院院长宋继强从检测、治疗和预防三个方面对如何利用AI技术抗疫做了介绍。在检测方面，英特尔帮助实现使用AI技术进行肺部CT初步检测，在疫情初期大幅提高了检测效率；在治疗方面，通过机器人进行医学护理降低了医

护人员的感染风险；在预防上，实现了通过人工智能进行检测，判断口罩生产是否合规，提高了生产效率，从多个层面共同助力了疫情防控。

滴滴智能控制首席科学家唐剑分享了滴滴出行在疫情期间利用 AI 技术赋能海外抗疫做出的努力。滴滴通过 AI 算法识别司机是否佩戴口罩，这是继免费向抗“疫”科研机构和企业开放 GPU 算力后，滴滴技术助力抗疫的又一举措，极大赋能了海外市场。

公众环境研究中心副主任王晶晶介绍了公众环境研究中心对支持疫情防控的努力。疫情爆发后，中心工作人员在短时间内对“蔚蓝地图”进行了升级调整，成为国内第一个将疫情情况精确到地级市的地图应用，对减轻疫情初期的公众恐慌起到了至关重要的作用。

杭州市萧山区数据资源管理局副局长孔秋实分享了杭州市萧山区的人工智能抗疫经验。萧山区政府在疫情爆发后第一时间组建了数据专班，研发部署疫情防控智能应用，推出了智能语音机器人、“萧山战疫”数字驾驶舱、“疫战”APP、萧山房东联盟等一系列应用，用人工智能与大数据赋能疫情防控一线，实现了政府决策科学化、调查摸排高效化和疫情处置精准化。

第二场专题对话主题为“后疫情时代的人工智能国际治理展望”，由清华大学计算机系人工智能实验室副主任张敏主持。

旷视科技副总裁蒋燕首先分享了旷视在抗击疫情上的主要贡献，她指出旷视在疫情早期成立了紧急研发攻关团队，短时间研发出人工智能测温系统解决方案，并第一时间投入到医院、交通枢纽、商超等人流密集的地区，有效地缓解了疫情的冲击，避免了由于人员聚集带来的病毒传播风险。

和玉资本管理合伙人曾玉从风险投资者和风险投资行业的角度谈论了疫情期间的变化。回顾疫情期间，在线办公、在线教育协同、在线医疗、以及远程问诊等行业迅速发展，这些背后都离不开人工智能技术的突破。同时，借助 AI 等技术手段，能够使企业、NGO 以及政府间形成有效合力，共同助力了疫情的防控与狙击。

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通过“日本疫情的教训”对后疫情时代的人工智能国际治理进行了展望。他认为围绕 AI 数码治理社会至少需要解决几个问题：第一，数据拥有权的归属以及应该如何被使用与分享，以及由谁为数据滥用负责。第二，个人隐私和个人的安全如何得到保护；第三，应该如何平衡国家、平台公司及社会间的关系；

第四，在目前国际政治使互联网走向碎片化的背景下，应如何维护互联网的整体性。

剑桥大学工程系高级研究员、图灵研究所人工智能项目主任阿德里安·韦勒认为，人工智能技术不应仅仅追求即刻的优化，同时也要思考系统怎样能够更加强健。在疫情下卫生健康方面，我们应开始关注怎样思考实现个人和集体扶持间的平衡，如应该如何分享健康数据，从而使得每个人都从中受益；怎样合理使用数字追踪技术进行疫情防控，这些不只需要某个国家的努力，更需要国际间确立相关的原则与规范。

加拿大国际治理创新中心总裁罗欣顿·麦德拉提出了疫情带来的两个思考。首先，疫情突出了当前的数字鸿沟，在我们已习惯线上交流的当下，世界上仍有二分之一的人没有任何网络接入，产生严重的数字鸿沟，随之也产生了网络安全以及供应链韧性的问题。此外，疫情扩大了人工智能技术在广泛使用中出现的各种问题。如在使用跟踪应用程序作为抗击流行病措施的情景下，如何搜集、存储和使用数据，以及从数据到其他治理问题的滑坡，包括隐私、安全和人权问题等都是目前亟需解决的难题。

2. 分论坛二：人工智能治理促进可持续发展

2020年12月18日下午，由联合国开发计划署承办的分论坛“人工智能治理促进可持续发展”成功举办。

联合国儿童基金会创新主任托马斯·戴文认为，如果用正确的方式引导，人工智能将潜力无限。他阐述了人工智能带来的两大挑战，首先是专业能力方面的差距，如今还没有能充分发掘人工智能潜力的专业技术，此外，人工智能方面的人才往往集中在发达国家；其次是技术方面的差距，世界上仍有将近一半的人还未联网。目前人工智能算法所使用的数据大多来自发达国家，这意味着它们几乎是在“茧房”中运行，嵌入具有偏向性的设定并主要解决发达国家的需求。为了应对这些挑战，托马斯提出两个可行的解决方案，第一是将人工智能的使用民主化，对中低收入国家人工智能的专业能力和专业技术建设进行投资。第二是建立治理体系，让第三方评估机构了解算法如何运作，从而保证人工智能的应用是不倾斜于任何人群的。他介绍了联合国儿童基金会在推进有关解决方案方面的工作，包括发布关于儿童人工智能的政策指导，以及投资能够让10亿弱势群体受益的科技企业。

麻省理工学院物理学教授马克斯·泰格马克阐述了人工智能国际间合作达成的

两个必要前提：共同的积极愿景和建立清晰的“红线”。他认为人工智能不是一个零和游戏。正如拥有一个共同的积极愿景能让我们即便在 20 世纪 80 年代的地缘政治紧张时期也能消灭天花一样，可持续发展目标是我们需要努力实现的另一个共同的积极愿景。泰格马克教授与其合作者的一项研究发现，人工智能有可能促成实现 79.3% 与社会相关的可持续发展目标，92.6% 与环境相关的可持续发展目标，以及 63.3% 与经济相关的可持续发展目标。与此同时，它也可能干扰剩余 37.8%、29.5% 和 31.7% 的以上对应领域的可持续发展目标的实现。因此，人工智能专家可以向生物学家和化学家学习，为人工智能的应用设定明确的“红线”，并引导人工智能走向人人皆可受益的方向。中国作为世界领先的科技大国，国际影响力不断增强，基于此，泰格马克教授进一步指出中国在促成人工智能国际合作方面具有得天独厚的优势。

北京智源人工智能研究院人工智能伦理与可持续发展中心主任、中科院自动化所中英人工智能伦理与治理中心主任曾毅教授基于一个对 AI4SDGs 智库所收集的研究案例的分析表示，在所有人工智能相关的出版物中，只有 0.1% 的出版物对可持续发展目标有直接贡献。其中，大部分案例与健康教育的可持续发展目标有关，却很少关注与环境相关的可持续发展目标。他分享了三个关于人工智能对社会影响的案例：案例一，为了规范学生的行为，教室里出现了人工智能监控，虽然这令人担忧，但却可有效防止校园欺凌的发生；案例二，人工智能可以用来监测和分析生物多样性数据，更好地保护野生动物，提高动物养殖福利；案例三，面部识别能定位接触者，有助于防疫工作，但也引发了与隐私相关的问题。他提出了一个有趣的观点，虽然东西方社会都将人工智能视为工具，但东方国家，如中国和日本往往将人工智能视为合作伙伴，而西方的多数国家更倾向将其视为竞争对手。他认为在未来的可持续的共生社会中，人类、动物、其他生物和环境终将和谐共存。

专题讨论环节由联合国开发计划署助理驻华代表张薇主持，北欧人工智能研究所执行董事克里斯蒂安·古特曼、清华大学公共管理学院副院长及清华大学全球可持续发展研究院执行院长朱旭峰、斯里兰卡技术基础设施智库主席罗罕·萨马拉吉瓦、联合国全球脉行动数据隐私专家米拉·罗曼诺夫、瑞典大使馆创新参赞南南·伦丁、以及地平线首席战略官郑治泰共六位座谈专家参与。首先对人工智能是推动还是阻碍可持续发展目标进行了讨论，列举了许多人工智能起促进作用的例子，如预测洪水、监测粮食安全、评估性别差距等。为了使人工智能应用更加广泛，需要建

立公众对人工智能的信任。此外，由于一些人工智能算法不需要大量的数据集，数据可用性差距这一问题也可以得到解决。专家们还对人工智能的治理架构提出设想，该框架在包容文化与社会差异的同时，也必须具有实用性和相关性，能够为商业领袖们提供投资的空间。关于如何平衡人工智能的监管和创新，专家们指出应划出清晰的“红线”，使用“沙盒方法”（The sandbox approach）来测试新的应用。人工智能创新和相关政策的出台应在可持续发展目标下朝着同一方向共同前进。

3. 分论坛三：人工智能与国际安全：挑战与机遇

2020 年 12 月 18 日晚上，由清华大学战略与安全研究中心承办的“人工智能与国际安全：挑战与机遇”分论坛成功举办。此次分论坛发布了中心与布鲁金斯学会联合研究项目“人工智能与国际安全”的初步成果，这个项目得到博古睿研究院和明德路基金会的共同支持。

布鲁金斯学会主席约翰·艾伦在主旨发言中提到，中美在人工智能等新兴技术领域的战略竞争日益明显，围绕着“人工智能军备竞赛”的说法不绝于耳。为了避免不可控的军备竞赛，他希望“人工智能与国际安全”项目中所讨论的安全关切，能够在政府层面对话和非官方论坛中得到进一步探讨。特别是针对人工智能所构成的四个领域的挑战和风险，包括人工智能技术的限制、人工智能导致冲突升级的风险、人工智能技术扩散的风险以及人道主义方面的风险，应当通过探讨从而降低这些风险发生的可能性。他认为现在正是规范技术发展和应用的关键期，呼吁美中两国的政府和专家团体能够抓住当下的机遇，围绕降低人工智能可能构成的国家安全风险的实际步骤达成共识。

清华大学战略与安全研究中心主任、人工智能国际治理研究院名誉院长傅莹在主旨发言中重点介绍了中国关于人工智能治理国际合作的主张，她说，今年 11 月中国国家主席习近平在二十国集团（G20）峰会上强调，中方支持围绕人工智能加强对话，倡议适时召开专题会议，推动落实 G20 人工智能原则，引领全球人工智能的健康发展。今年 9 月，中国国务委员兼外长王毅谈到《全球数据安全倡议》，表示希望国际社会在普遍参与的基础上就人工智能的安全问题达成国际协议，支持通过双边或多边协议的形式确认倡议中的承诺。他提出有效应对数据安全风险应遵循的三项原则，包括秉持多边主义、兼顾安全发展和坚守公平正义等。傅莹提到，中美关于人工智能安全治理的联合研究项目是一个成功的案例，双方都从中受益良

多，研究结果也有很好的参考价值。既然人工智能武器化不可避免，我们需要寻找合适的治理路径。人类必须汲取历史教训，例如对核武器的治理共识就发生的太晚，导致人类曾经面临巨大威胁，互联网的治理也没有能够在早期实现。希望此次面对人工智能，尤其是智能武器的治理，人类能走在技术变革的前面，充分认识其风险，早些达成治理共识。在讨论人工智能国际治理时，人们经常引用在核武器治理方面享有声誉的帕格沃什科学和世界事务会议 Pugwash 机构（帕格沃什科学和世界事务会议），这一机构在核军控领域起到了积极作用。在推动人工智能的国际治理方面，也可以参考这样的方式，成立一个由科研人员、政策制定者和政策建议者组成的“人工智能国际治理委员会”。

在座谈环节，专家就中美如何在人工智能治理领域合作展开深入讨论并表示，中美两国是人工智能大国，都有防范军用人工智能技术风险和完善治理的需求。人工智能技术的滥用或者偏差有可能冲击战略稳定、改变交战规则，抬高军事风险。中美都有较强的风险管控需求和共同利益。因此，中美在人工智能国际治理中有很大的合作空间。有专家指出，人工智能在数据收集、算法训练和战场应用中，必须遵守国际法中的比例原则和区分原则，不能攻击平民或民用设施，造成过多的伤害。中方专家提出采用数据收集的“红绿灯”规则，禁止人工智能武器攻击民用目标，只能对完全确认的军用目标发动进攻。这个措施可以使人工智能更好地遵守国际法。最后，专家们表示，包括中美在内的世界各国可以通过政策交流、互换文件、学者探讨等多种方式，增信释疑，实现在人工智能安全问题上形成国际治理共识。

4. 分论坛四：人工智能治理国际合作

2020年12月19日上午，由联合国开发计划署承办的分论坛“人工智能治理国际合作”成功举办。

洛克菲勒基金会高级副总裁齐亚·汗首先指出，虽然有许多人工智能伦理准则已公布，但很少有政府去实施具体的措施。此外，仅靠伦理协议还不足以保证“好的人工智能”的发展，我们需要为人工智能量身定制规则。他分享了几点建议，一是每个人对人工智能的理解不尽相同，我们需要注意如何在不同的社会政治环境中接纳这些差异；二是许多跨国公司正开发利用人工智能技术，但大多数监管框架是局限在国家层面的；三是在大流行病和战争等危及生命的情况下，我们需要对人工智能治理尽快达成共识；四是有关机构需要对人工智能应用的安全性、有效性和公

平性进行评估。

新加坡信息通信媒体发展管理局（IMDA）助理首席执行官杨子健认为，国际人工智能治理框架应该具有前瞻性、开放性和共同操作性，同时具有商业合理性、实用性和现实性。他阐述了新加坡在人工智能治理方面所做的努力包括两大“支柱”，一是保护个人数据的法律，二是鼓励信任人工智能发展的工具，如“可信数据共享框架（trusted data-sharing framework）”“数据监管沙盒（data regulatory sandbox）”等。为促进负责任的人工智能的发展，新加坡已经实施了一套人工智能治理举措，包括自我评估指南和案例研究。另外，新加坡已经开始为人工智能治理领域的专业人士提供培训和认证项目。他指出只有通过合作，以上所有的这些努力才能实现，今后将以“以人为本”作为指导方针，积极审视人工智能应用成果以确保其关联性。

专题讨论环节由联合国开发计划署驻华副代表戴文德主持，韩国延世大学未来政府研究所所长文杰，印度 Omdena 创始人鲁德拉得·米特拉、马来西亚数字经济发展局首席执行官苏里纳·舒克里、香港中文大学数据科学与政策研究所所长黄伟豪、未来社会 AI 倡议联合创始人兼主任赛勒斯·霍德斯、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、以及南非 Webber Wentzel 律师事务所首席信息官沃伦·海罗共六位专家参与，就人工智能国际合作和治理提出了两套“相对”的方案——“被动/主动”方案和“自上而下/自下而上”方案。专家们一致认为，需要在“被动式”和“主动式”做法之间取得平衡，以鼓励人工智能的创新，同时减少风险。一些专家主张采取“自下而上”的方法，因为大多数人工智能的伦理规则缺乏来自民众的声音。部分专家担心虽然人工智能技术在应对全球挑战方面具有巨大潜力，但大多数创新都是由经济利益推动的。讨论到数字主权的话题，几位专家一致认为，我们需要更全面地思考负责任的人工智能准则问题，而不是维护各自利益。

5. 分论坛五：人工智能与数据安全

2020年12月19日上午，由北京智源人工智能研究院承办的“人工智能与数据安全”分论坛成功举行。分论坛由北京智源人工智能研究院人工智能伦理与可持续发展中心主任、中科院自动化所中英人工智能伦理与治理中心主任曾毅主持。

剑桥大学智能未来研究中心研究员丹尼特·加尔在主旨演讲中区分了人工智能

安全的两大重要概念，即 Safety 和 Security 的概念。Safety 更偏向内部，针对的是一些无伤害的预防措施，从系统中保护环境。Security 更倾向于对外部攻击的防御，聚焦在那些有意对个人、组织和财产施加侵害的行为，从环境中保护系统。她认为，科技影响着所有人，它的安全和保障也是如此。然而，我们应对这些影响的能力远远落后于技术的开发和使用。人工智能和数据流进一步复杂化了这一动态，引入并连接了新的机会和新的漏洞。丹尼特·加尔建议，应加强交流、推动合作，以减少脆弱性，在全球范围内倡导并激励数据安全治理工作。

荷兰埃因霍温科技大学教授文森特·穆勒在主旨演讲中，分成两步介绍了超级智能和通用人工智能的长期风险，即从超级智能到技术奇点，再从技术奇点到生存风险。超级智能指的是在许多非常普遍的认知领域远远超过当今人类最佳智力的人工智能。从超级智能到技术奇点，他强调三个前提假设，分别是我们拥有更快的速度和更多的数据，不需要认知科学，以及十分紧急。从技术奇点到生存风险，他强调了两点前提假设，理性主体决策理论适用于人工智能，以及智能和最终目标的正交性。

美国路易斯维尔大学教授罗曼·扬波尔斯基在主旨演讲中，分享了关于人工智能、安全和防御的看法。当前超级智能正在来临，超级智能意味着超级聪明、超级复杂、超级快速、超级控制、超级病毒。他谈到关于人工智能的担心，重点介绍了最近的几项研究工作，包括通往危险的人工智能的路径分类学、人工智能安全不可能的结果、缓解负面影响、人工智能的限制问题等。他认为，人工智能失败的时间轴呈指数趋势，人工智能失败的频率和严重程度都会增加，这与人工智能的能力成比例。

在专题讨论环节，奇虎 360 人工智能研究院院长邓亚峰、数据科学家麻丹彤、腾讯研究院高级研究员曹建峰、北京师范大学网络法治国际中心执行主任吴沈括分别就人工智能与数据安全面临的机遇、挑战和发展趋势等发表了一系列前瞻洞见。

曹建峰分享了对算法偏差和深度合成（deep synthesis）的看法，并提出了解决问题的建议。当前人工智能系统必须使用大量数据训练算法。因此，人工智能开发应该对数据质量有合理的控制。但数据并不是偏见的唯一来源，解决算法偏见还需要 AI 从业人员秉持伦理价值。此外，人工智能开发与隐私保护直接存在紧张关系，为解决这些问题，曹建峰给出了四点建议，人工智能与数据产品开发需要遵循“通过设计的伦理”（ethics by design）理念，将公平、安全等伦理价值融入

产品设计全周期，并有防止滥用的保障措施；二是在企业、国家乃至国际层面建立负责任的人工智能开发和数据使用的道德准则；三是在国际社会分享最佳实践经验；四是业界需要探索开发隐私保护型的 AI 技术方法，如联邦学习等，来在隐私保护与数据利用之间实现更好平衡。

吴沈括认为国际社会需要面对新技术环境下数据违法行为带来的新挑战和法律漏洞。首先，由于不同地区、国家法律法规的缺失和差异，普遍导致对新型数据违法行为司法规制不足。其次，各国现行法律制度框架尚不能充分保障跨境证据收集与交换的快速高效。最后，在网络匿名的背景下，一些国家的监测、预警和应急机制越来越难以有效应对脆弱性问题的快速增长。因此，在全球化深入发展的新时代，数据违法行为的跨国衍变要求更高水平的国际合作治理，建立信任世界。

邓亚峰总结了人工智能在数据安全方面存在的潜在威胁。首先，由于人工智能是快速出现的新应用，人工智能应用生产和部署的各个环节都暴露出了一系列的安全风险。其次，人工智能和大数据应用往往与大量隐私数据关联，这些数据往往需要被送到云端进行分析和计算，造成了比原来更严重的数据泄露和安全风险。第三，虽然从技术角度看，人工智能体很长一段时间都无法拥有真正的自我意识，但是，我们无法排除一些不法分子会对人工智能加以利用，而这会带来大量的社会伦理问题。最后，由于人工智能在短期内会造成一部分从事简单重复性工作的人类失业，这会引发一定的恐慌和排斥。此外，人工智能也因为技术的不成熟，会被认为在识别结果上存在一定的歧视。他认为，人工智能和数据安全是全世界都统一面临的问题，需要全世界的组织和个人联合起来一起行动。

五、第二届清华大学人工智能合作与治理国际论坛

第二届清华大学人工智能合作与治理国际论坛将于 2021 年年底举行，会议信息即将发布，敬请关注。

清华大学人工智能合作与治理国际论坛网址：<https://www.tsinghuaforum.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