《2019 全球人工智能治理年度观察》主要内容及启示

李 辉，谢旻希，王迎春
（上海市科学学研究所，上海 200235）

摘 要：基于《2019 全球人工智能治理年度观察》，本文分析了报告中来自全球的 44 家代表性机构对 2019 年人工智能治理进展的见解，并提出当前“全球人工智能治理体系”正在孕育形成的观点。本文对当时全球人工智能治理体系孕育期的发展趋势和特点也进行了研究，发现人工智能治理的国家共同体正在形成，治理规则正在从“软法”向“硬法”转换，企业成为微观治理的主导力量。最后结合这些趋势和特点以及我国人工智能治理现状，对我国参与并进一步推进全球人工智能治理体系建设提出若干建议。

关键词：人工智能；全球治理；治理体系；国际共识
中图分类号：F124.3；G321 **文献标识码**：A DOI：10.3772/j.issn.1009-8623.2021.04.006

人工智能的快速发展，正在对政治、经济、社会以及国际关系产生重大影响，既有正面的促进作用，也有负面的潜在风险。因此，人工智能治理已成为当前全球政治、产业和学术等领域的焦点议题。随着越来越多利益相关方的参与以及政策文本和研究成果的发布，人们迫切需要从越来越繁杂的文献中，准确把握全球人工智能治理的发展态势和主要特点，从而为相应的政策制定和学术研究提供基础。

上海市科学学研究所于 2020 年 4 月发布的《2019 全球人工智能治理年度观察》[1]（下文简称《2019 观察》）是在此方面的一次尝试。该报告以“2019 年度全球人工智能治理进展”为主题，邀请全球政产学研界具有重要话语权的代表人士对 2019 年全球人工智能治理的重要进展进行回顾，图灵奖获得者、美国人工智能年会（AAAI）主席、欧洲议会议员以及联合国相关组织代表等 44 组专家（50 位）提供了评论文章 [2]。

由于人工智能治理问题的复杂性，实际上全球同行也都迫切需要获取其他背景（专业、行业、领域、国家等）的专家对这一问题的看法 [3]。该报告呼应了这一需求，发布之后在国际人工智能治理领域产生了广泛影响。业内较为知名的蒙特利尔人工智能治理研究所评论认为：《2019 观察》所包含的 44 篇文章，每一篇都评价了 2019 年人工智能治理领域的重大事件。除了包含发人深省的见解外，该报告还提供了一个供全球人工智能治理专家、众多研究中心、智库和相关组织交流的优质平台 [4]。从某种程度来看，这份报告为全球人工智能治理现状提供了一种“思想众筹”，既为参与的作者提供了表达“我们在做什么”的机会，也为全球同行提供了看见“别人在做什么”的平台。这些专家评论了这一领域的趋势和特点，而从这些代表性的观点中，又可以总结出更具统领性的趋势和特点。

以报告英文标题搜索国际知名的人工智能社区以及领英、推特、脸书等媒体，可以看到大量国际人工智能治理机构都对这一报告进行了推

第一作者简介：李辉（1981—），男，副研究员，主要研究方向为人工智能治理、科技与社会。
收稿日期：2021-02-26

介和评论。其中具有代表性的一种评论是，这份报告是新冠肺炎全球流行之前，国际同行对人工智能治理的一次总结和展望。当然，即将发布的《2020 全球人工智能治理年度观察》，将提供疫情暴发状态下的全球人工智能治理进展。本文主要针对《2019 观察》中 44 组专家的观点进行了凝练总结，力图总结出疫情暴发前的全球人工智能治理态势和特点，并最终为中国下一步的人工智能治理工作提供相关建议。《2020 全球人工智能治理年度观察》体现了全球在疫情冲击下的人工智能治理进展，我们将另文专论[①]。

1　关于全球人工智能治理现状的 44 种评论

每一位专家都基于自己的专业、行业、领域背景，对全球人工智能治理重要进展做出判断。该报告分为六部分：人工智能科学家的思考与倡议、人文社会科学专家的关注与进展、产业界的实践和探索、国际组织相关政策进展、国家和地区相关政策进展以及来自中国的声音。以下对这 6 种背景的专家观点逐一进行提炼说明。

1.1　人工智能科学家的思考与倡议

人工智能有可能造成诸多负面影响，因此开发人工智能系统的技术专家，是当然的被治理对象。科学家群体自身也提供治理方案。该部分报告有 5 家单位的科学家参与，主要提供了两种解决方案。一种是用新技术解决原有技术带来的问题。其中 Stuart Russell 教授多年来一直努力推动“开发有益的人工智能（AI）系统”，在国际人工智能界有广泛影响。香港科技大学的杨强教授呼吁的“联邦学习”，就是希望用新的技术来解决人工智能带来的负面影响。另一种方案，就是用制度对科学家的研究进行规范，如香港科技大学的冯雁教授提出了人工智能算法发表的规范性问题。5 组技术专家的核心观点如表 1 所示[②]。

1.2　人文社会科学专家的关注与进展

人文社会科学是提供治理规则的思想源。法学、管理学、国际关系、哲学等人文社科领域专家纷纷涉足人工智能治理领域。参与报告的 12 家单位的人文社科专家，专业背景较为多元。他们讨论了在规制人工智能方面的进展，比如，牛津大学未来人类研究所人工智能治理中心主任 Allan Dafoe 罗列了 2019 年新成立的多家人工智能治理研究机构，以此表明人工智能治理已经成为一门显学。在研究内容方面，德国美因茨约翰内斯古

表 1　人工智能科学家的思考和倡议

专家	主题	观点
John Hopcroft（美国） 康奈尔大学计算机科学系工程与应用数学教授，图灵奖获得者	《迎接科学革命，培育更多人才》	当前的人工智能并没有科学基础上的革命，所以容易出现从“偏见”到“偏见”的问题
Stuart Russell（美国） 加州大学伯克利分校人类兼容人工智能中心主任，倡议发展友善及安全人工智能的先驱之一	《开发有益的人工智能系统》	科学家应当开发有益的人工智能系统
杨强（中国香港） 微众银行首席人工智能官，香港科技大学计算机科学和工程系主任、教授	《联邦学习的重要性》	提倡从技术角度，如开发联邦学习，来解决当前人工智能治理问题中比较突出的隐私问题
冯雁（中国香港） 香港科技大学教授，香港科技大学人工智能研究中心主任	《为开发符合伦理的人工智能建立正式流程》	应当针对当前的人工智能技术路线设置规范的算法审查制度
Roman Yampolskiy（美国） 路易斯维尔大学网络安全实验室主任	《人工智能治理和人工智能安全》	不能仅仅讨论伦理问题，更要关注人工智能系统本身的安全问题

① 本文第一第二作者为《全球人工智能治理年度观察》系列报告执行编辑。

② 每组专家可能不止一人，表格中仅列出第一作者，下同。

腾堡大学的 Petra Ahrweiler 教授和清华大学的苏竣教授都谈到了如何评估人工智能带来的社会影响。越来越多研究机构的成立，越来越多治理方案的推出，意味着人工智能治理的思想源越来越丰富，这些治理方案都会成为未来全球人工智能治理的基础性思想资源。12 组人文社科专家的核心观点如表 2 所示。

1.3　产业界的实践和探索

企业是技术的应用主体，因此从某种程度来说，是被治理的对象。但是对于负责任的企业来

表 2　人文社会科学专家的关注与进展

专家	主题	观点
Allan Dafoe（英国） 牛津大学未来人类研究所人工智能治理中心主任	《人工智能治理领域的迅速发展》	2019 年新建立了若干人工智能治理研究机构，从中可以看出专业的人工智能治理研究正在快速建制化
Gillian Hadfield（加拿大） 多伦多大学施瓦兹·赖斯曼技术与社会研究所创始所长、首席教授	《人工智能与人类价值观的有效统一：从“应该”到“如何”》	多伦多大学刚刚成立一个新的人工智能治理研究所，该研究所的使命是，把人工智能的社会影响从有些泛滥的“应该”向如何落地的“如何”转移
苏竣（中国） 清华大学公共管理学院教授，教育部长江学者特聘教授	《采用长周期、多学科的社会实验方法研究人工智能的社会影响》	苏竣教授发起了一个有着丰富内容的“社会实验”计划，通过这一计划对人工智能治理进行全面深入的研究
Thilo Hagendorff（德国） 德国图宾根大学科学与人文伦理国际中心副教授	《将人工智能原则由软法转化为硬法》	经过 2019 年的发展，各界迫切呼吁人工智能的治理应从“软法”向“硬法”过渡
Petra Ahrweiler（德国） 德国美因茨大学教授，欧洲技术与创新评估研究院院长	《用跨学科方法探索人工智能治理研究》	多学科交叉展开人工智能评估的探索
Williams Robin（英国） 爱丁堡大学科学技术与创新研究所所长	《对人工智能预期治理的看法》	提出了预期治理的概念
Colin Allen（美国） 匹兹堡大学科学哲学和科学史系特聘教授	《媒体宣传需要深度理解技术》	新闻媒体的关注正在影响人工智能治理
Poon King Wang（新加坡） 新加坡科技设计大学李光耀创新型城市中心主任	《未来的工作：以“任务”为基本单元的研究分析——新加坡的探索》	用定量的方法研究了人工智能对新加坡就业的影响
Ferran Jarabo Carbonell（西班牙） 赫罗纳宗教科学研究所教授	《发展为人类服务的人工智能》	从宗教关怀的视角，强调了人工智能服务于人的必要性
王小红（中国） 西安交通大学哲学系教授，人工智能伦理研究专家	《在增进文化共识基础上加强人工智能治理全球协作》	未来的人工智能治理更应该增进各种文化之间的共识
杨庆峰（中国） 复旦大学应用伦理研究中心教授	《人工智能治理的三种模式》	梳理了人工智能治理的三种理论基础

说，这也是引领人工智能发展的好机会。中国的人工智能独角兽企业旷视科技以及公益性的社会组织 OPAI 参与了本次报告。略有遗憾的是，脸书、谷歌、特斯拉等全球人工智能巨头企业没有参与。在法律伦理规则没有完全成型的背景下，企业通过市场行为，正在探索具备可操作性的治理措施 [5]。2019 年尤其值得关注的是，一些企业开始发展针对假新闻、假视频的检测技术。产业

界的投入让人工智能治理的可操作性变强。另外，2019 年讨论最为激烈的一个话题是 OpenAI 当年在发表 GPT–2 的过程中引发的争论，在《2019 观察》中有两位专家给予了评论。7 组产业界专家的核心观点如表 3 所示。

1.4 国际组织相关政策进展

人工智能治理是事关全人类与人工智能技术关系的问题，具有天然的全球性，也因此成为国际关系的重要议题。联合国相关机构、经济合作与发展组织（OECD）、世界经济论坛、世界银行等政府间组织都在大力推进人工智能全球治理体系。本部分主要有 5 位专家参与，如表 4 所示。其中值得关注的是，有 3 位专家重点谈及了经济合作与发展组织在 2019 年发布的人工智能治理原则。

表 3 产业界的实践和探索

专家	主题	观点
印奇（中国） 旷视科技联合创始人兼首席执行官	《直面人工智能治理挑战，企业要有所作为》	从一个中国企业角度讨论了对人工智能治理的最新思考和探索
Don Wright（美国） 美国电气电气工程师学会（IEEE）标准化协会前主席	《可信赖人工智能和公司治理》	介绍了美国电气电子工程师学会进一步针对人工智能产业应用的治理规则制定情况
Miles Brundage 等（美国） Open AI 人工智能政策团队	《2019：推动负责任发表规范的一年》	开发者在发布算法时应谨慎，Open AI 的探索
Seán ÓhÉigeartaigh（英国） 剑桥大学未来智能研究中心和生存风险研究中心主任	《可能用于恶意用途的人工智能研究：发表规范和治理方面的考虑》	算法审查制度事实上是科学社群的关注焦点，对 2019 年 Open AI 在发表 GPT-2 的过程中引发的争论进行了回应
Helen Toner（美国） 美国乔治城大学安全与新兴技术中心战略主任	《GPT-2 开启了人工智能研究社区对于发表规范的讨论》	同样对 2019 年 Open AI 在发表 GPT-2 的过程中引发的争论进行了回应
Millie Liu（美国） 第一星风险投资公司投资人	《企业人工智能应用中的伦理挑战——来自产业界的观察》	分析了当前企业在实际的人工智能应用时的伦理挑战
Steven Hoffman（美国） “创始人空间”首席执行官，硅谷孵化器和投资人	《人工智能治理，政策制定者应利用市场力量》	企业不仅仅是人工智能治理的对象，它们也更有可能发挥力量来解决人工智能伦理问题

表 4 国际组织相关政策进展

专家	主题	观点
Irakli Beridze（荷兰） 联合国人工智能和机器人中心主任	《掌握人工智能治理的双刃剑》	既要看到人工智能带来的伦理问题，也要看到它对一些全球性难题的正面作用
Wendell Wallach（美国） 耶鲁大学教授，全球知名科技伦理专家	《寻求灵活、合作和全面的人工智能治理国际机制》	全球人工智能治理应该寻求敏捷治理方案
Cyrus Hodes（法国） 阿联酋总理府人工智能部前部长顾问、经济合作与发展组织人工智能专家组成员	《国际社会人工智能治理意识的觉醒》	大量国际组织在提出人工智能治理原则，经济合作与发展组织工作成效明显
Nicolas Miailhe（法国） “未来社会”创始人、经济合作与发展组织人工智能专家组成员	《人工智能治理，从原则到实践的转变》	经济合作与发展组织在人工智能的治理方面做出了很大的进步，全球化共识在逼近

续表

专家	主题	观点
Jessica Cussins Newman（美国） 加州大学伯克利分校、未来生命研究所研究员	《OECD 人工智能原则——人工智能治理的全球参考》	《OECD 人工智能原则》的基本内容介绍
陈定定（中国） 暨南大学教授，海国图智研究院院长	《人工智能治理成为国际关系的重要议题》	讨论了国际关系界对人工智能治理的关注

1.5　重要国家和地区相关政策进展

和所有新兴技术的治理一样，国家是人工智能治理的当然主体。欧盟、英国、日本、新加坡等地区和国家，是当前制定人工智能治理规则的引领者。尤其是欧盟，在人工智能治理的规则制定中投入了大量的政策资源。2019 年欧盟发布了《可信赖人工智能伦理准则》，在全球有极大影响[6]。当然，国情不同，规则制定的方向也有所不同。新加坡发布的人工智能规则虽然影响不及欧盟，但可操作性却更强，对于小国有极大的参考价值。这一领域的 8 组观点如表 5 所示。

1.6　中国举措和声音

中国在人工智能技术开发和产业应用方面位于世界前列，但是治理方面的声音鲜少为全球同行所听到。7 位中国专家，包括前外交部副部长傅莹，科技部新一代人工智能发展研究中心主任赵志耘等，如表 6 所示，分别从国家关系、科技发展、工信发展、国家标准、地方探索以及学术研究等方面，多角度

表 5　国家和地区相关政策进展

专家	主题	观点
Eva Kaili（希腊） 欧洲议会议员	《欧洲议会应对人工智能治理的价值理念》	介绍了欧洲议会关于人工智能治理的工作理念和未来打算
Francesca Rossi（意大利） IBM 人工智能伦理全球负责人，IBM 研究院特级研究员、欧盟人工智能高级别专家组成员	《多边方法的典范——欧盟人工智能高级别专家组》	介绍了《可信赖人工智能伦理准则》中欧盟人工智能高级别专家组的工作理念和内容
Charlotte Stix（荷兰） 欧盟人工智能高级别专家组协调人，《欧洲人工智能新闻》主编	《欧盟采取“可信赖人工智能”的执行路线》	对欧盟《可信赖人工智能伦理准则》制定中的工作模式进行了介绍
Angela Daly（英国） 斯特拉斯克莱德大学法学院（苏格兰）副教授	《英国人工智能伦理的驱动力》	介绍了英国政府人工智能治理的工作机制，尤其介绍了专门治理机构数据伦理与创新中心的作用
Danit Gal（以色列） 联合国秘书长数字合作高级别小组技术顾问	《东亚人工智能伦理和治理政策地方化》	认为东亚在人工智能伦理和治理方面有着深刻的传统文化烙印
Arisa Ema（日本） 东京大学副教授，日本内阁以人为中心人工智能社会准则专家组成员	《日本人民对人工智能治理和伦理的担忧和期望》	介绍了日本人工智能治理最新情况，尤其提到了日本产业界开始积极介入人工智能治理工作
Goh Yihan（新加坡） 新加坡管理大学法学院院长、教授、人工智能和数据治理中心主任	《新加坡人工智能伦理和治理举措》	介绍了新加坡发布的亚洲首个人工智能治理框架等工作
Urvashi Aneja（印度） 印度智库 Tandem Research 联合创始人兼主任	《印度在人工智能时代面临的重大挑战：不平等和增长之间的矛盾管理》	对印度在人工智能治理方面的工作提出了更高的期望

表 6　来自中国的声音

专家	主题	观点
傅莹（中国） 清华大学战略与安全研究中心主任	《结伴同行，合作共赢》	在人工智能治理问题上，全球应当合作，中国和美国作为大国，也应该以合作为主
赵志耘（中国） 科技部新一代人工智能发展研究中心主任，中国科学技术信息研究所党委书记、研究员	《中国人工智能治理取得积极进展》	介绍了中国在人工智能治理方面的基本理念以及在 2019 年的主要工作和积极进展
李修全（中国） 中国科学技术发展战略研究院研究员	《从治理原则走向细化落地，更加需要多方参与、协同治理》	治理原则细化落地应注重协同包容，关注弱势群体，注重多元参与和对话沟通
段伟文（中国） 中国社科院科技和社会研究中心主任、教授	《中国走向稳健敏捷的人工智能伦理和治理框架》	介绍了中国在人工智能治理和数据治理方面的主要进展
栾群（中国） 工信部赛迪研究院政策所所长	《全球化与合伦理成为人工智能治理共识——中国产业界的伦理关注》	介绍了中国人工智能产业发展中的伦理治理进展
郭锐（中国） 中国人民大学副教授、国标委人工智能社会伦理项目负责人	《人工智能治理的造福于人和可问责性原则——在中国人工智能标准化制定中的理念》	介绍了在国家标准制定中秉承的理念
王迎春（中国） 上海市科学学研究所科技与社会研究室主任	《推动人工智能让城市和生活更美好》	介绍了上海国家人工智能创新发展试验区的情况和世界人工智能大会（上海）治理论坛的情况

反映了中国在人工智能治理领域的积极进展。

2　结论：全球人工智能治理体系处在形成关键期

人工智能技术是全球通用技术，必然需要全球治理体系。全球人工智能治理仍处于探索阶段[7]。基于对全球 44 家机构对 2019 年全球人工智能治理进展评论的分析，本文得出，全球人工智能治理体系正在孕育形成，呈现出治理主体多元性、治理措施多样化等特点。具体而言，国际上各类治理主体，包括国际组织、各国政府、行业组织、企业、技术社群、社会组织等，都积极推进相关研究，并争相提出治理主张。而治理规则呈现多样性特点，各类治理主体基于自身能力发布的人工智能治理方案包括：规范、规则、法律、标准、原则、倡议、宣言等。孕育期更重要的特点是，虽然不同治理方案之间存在博弈融合，但是各利益相关方对达成全球共识普遍抱有极大期待。可以预见，基于多方的期待以及各种治理手段的尝试，全球人工智能治理体系的孕育形成具备了良好的国际环境。基于《2019 观察》来看，现阶段全球人工智能治理有一些值得关注的发展趋势和特点。

2.1　人工智能治理的国家共同体正在形成

从全球来看，人工智能由“谁来治理”，首先的主体是国家以及国家之间的合作组织。不同国家在发布自己的治理规则的同时，也必须寻求与其他国家之间在规则上的通约。从全世界范围来看，作为政府间组织，经济合作与发展组织发布的原则获得了广泛的支持。《2019 观察》中三位参与者 Cyrus Hodes、Nicolas Miailhe 和 Jessica Cussins Newman 都提到了经济合作与发展组织，他们介绍了该组织通过一系列的组合拳，全力推动其规则的全球化。经济合作与发展组织的规则推广，很有可能成为未来人工智能治理推广的一个范本。其具体推广模式如下：

（1）发布原则并争取越来越多的国家支持。经济合作与发展组织于 2019 年 5 月 22 日公布了其《OECD 人工智能原则》，所有 36 个经济合作与发展组织成员国以及多个非成员国都签署了该原则。

此外，欧盟委员会也表示支持，该原则还获得了中国、俄罗斯等国的认可，是目前受到支持最多的人工智能原则。

（2）建立“人工智能政策观察站”。为了实施其人工智能原则，经济合作与发展组织还宣布建立一个“人工智能政策观察站”，该观察站将提供人工智能指标、政策和实践方面的证据和指导，并作为一个促进人工智能政策对话和分享最佳实践的中心。

（3）启动“人工智能全球伙伴关系”。法国和加拿大在 2019 年 8 月的七国集团会议上宣布启动由经济合作与发展组织主办的“人工智能全球伙伴关系”（GPAI）。“人工智能全球伙伴关系”的目标是将全球高水平的人工智能科学家和专家聚集在一起，促进合作伙伴在人工智能政策制定方面的国际合作与协调。

（4）打造“全球人工智能人类论坛”。作为“人工智能全球伙伴关系”的配套活动，法国于 2019 年 10 月底在巴黎主办了第一届“全球人工智能人类论坛”。

不过，经济合作与发展组织是西方国家为主的国际组织，因此具有先天的局限性。实际上联合国相关机构也正在推动相关规则的全球化。可以确定的是，寻求全球协议，将是全球人工智能治理体系建立的关键。

2.2　治理规则正在从“软法”向“硬法”转换

各种治理主体基于自身能力发布的人工智能治理方案包括规范、规则、法律、标准、原则、倡议、宣言等，目前以不具法律效力但是有方向指导意义的倡议、宣言、原则等居多。随着近几年各种组织机构不断发布人工智能准则，越来越多的研究者对纯粹宣言性质的准则提出了批评。一方面，宣言性质的准则对人工智能的健康发展没有实质性的约束，另一方面，一些企业或者组织以发布伦理准则作为宣传甚至“道德清洗”的策略。《2019 观察》中，德国图宾根大学科学与人文伦理国际中心副教授 Thilo Hagendorff 强调了这一呼声，即人工智能治理应该从“软法”向“硬法”过渡。欧盟人工智能伦理准则颁布后，有评论者认为，它可以增强消费者对欧洲发展人工智能的信心，从而为企业提供对抗硅谷和深圳竞争对手的解决方案 [8]。

《2019 观察》中一些专家的文章，已经展现出这样过渡的迹象。其中代表性的有，美国加利福尼亚州要求把所有试图影响加州居民投票或购买行为的自动在线账户公开标识为机器人；旧金山等地也出台法律禁止人脸识别应用 [9]。

当然，最根本性的硬法是技术本身的革新，这就需要技术专家开发出与人和谐的人工智能系统。蒙特利尔人工智能治理研究所也注意到了《2019 观察》中一些专家的提法，其发布的报告特别提及，“专家杨强教授还提到了联合学习、差分隐私和同态加密等新技术的重要性，以及它们在确保人工智能被用于造福人类方面的重要性。”[4]。

《2019 观察》中反映的从“软法”到“硬法”转化的态势，实际上在 2020 年得到了验证，比较重要的一个例证是，欧盟在 2019 年发布《可信赖人工智能治理规则》的基础上，于 2020 年进一步发布了《人工智能白皮书——通往卓越和信任的欧洲路径》。该白皮书综合提出了有关伦理的注意事项、法律义务和技术基础设施等，目的是把人工智能治理置于可操作的层面。

2.3　企业成为微观治理的主导力量

一般认为，人工智能相关企业和科学家是人工智能治理规则的主要应用对象。但是从最新的发展趋势来看，企业逐渐从被动接受治理规则，转变到主动谋求制定规则甚至开发具体治理工具。美国电气和电子工程师协会商业委员会（IEEE）在 2020 年第一季度发布一份题为《对企业使用人工智能的呼吁》的倡议，其中就提到企业应根据用户的伦理需求来研发产品。《2019 观察》中美国电气和电子工程师协会商业委员会标准化协会前主席 Don Wright 介绍了美国电气和电子工程师协会商业委员会发布这一倡议的背后逻辑：一方面，企业负责创新，而政府规则通常落后于创新发展，无法跟上企业的发展需求；另一方面，各种文化群体的用户伦理偏向并不一致，企业最前端接触用户，理解客户的伦理偏向，并且能够通过不断试错理解和培养用户群体的伦理观念。

另外，单纯依靠政府能力难以实现海量的微观层面的治理，企业开始发展具体的规范操作。来自硅谷的知名投资人 Steven Hoffman 在《2019 观察》中介绍了这方面的情况，如谷歌发布了一个数据集

来帮助检测合成声音，脸书、Partnership AI 和其他组织发起了“深度造假”视频检测比赛。市场力量让人工智能治理更具可操作性和可实现性。

3 中国进一步推动人工智能治理工作的若干启示

我国的人工智能治理体系正在孕育形成。参与《2019 观察》的中国专家，对中国 2019 年的人工智能治理工作有比较全面的介绍。结合最新的进展，我国已经有初具体系化的人工智能治理体系，如国家层面的科技伦理委员会、国家新一代人工智能治理专业委员会、国家新一代人工智能战略咨询委员会等机构，地方层面有科技部批准的北京、上海等 15 个人工智能创新发展试验区，有很多地区建立了自己专门的研究或咨询机构，智库高校科研院所也建立了众多人工智能治理研究机构。2019 年 7 月，科技部新一代人工智能治理专业委员会发布了八条治理准则。国家安全标准委员会、国家人工智能标准化总体组等也都发布了相关治理文件。还有个人信息保护法、人脸识别等方面的立法也已经在征求意见。

但是基于《2019 观察》，并结合我国也正在建立中国的人工智能治理体系，以及当前全球人工智能治理体系“孕育期”的发展趋势和特点，对于我国进一步推进人工智能治理体系建设以及参与全球人工智能治理体系建设，我们认为还可以有以下几点启示。

3.1 进一步推动人工智能风险评估

近些年，诸多组织发布了人工智能治理原则。但是由于原则没有约束力，因此对人工智能的负面影响没有实质上的规范，而推动实质规范的前提是需要对人工智能的负面影响进行评估。经济合作与发展组织的人工智能规则中，也强调“应该不断进行评估和管理”。

但是人工智能的风险评估不仅是制度问题，同时也是科学问题。随着技术发展的复杂性以及与社会交互的不确定性，新的技术评估需要能够精准、实时地判断和预测技术的社会综合影响和社会变革，因此需要相应方法支撑其更好地发挥作用。清华大学苏竣教授在《2019 观察》中提到了“社会实验”的方法，证明中国的人工智能风险评估已经迈出了可喜的一步。

当然，人工智能评估工作在国际上仍然属于探索性的前沿课题。我国由于疫情控制得当，人工智能产业继续稳步发展，因此有必要通过一定方式进一步推动人工智能风险评估工作。

3.2 打造对接渠道，推动中国专家参与外交讨论

由于人工智能治理是新兴国际议题，还缺乏有效的国际交流平台。《2019 观察》由中国智库发起，得到全球众多人工智能治理机构和知名专家的响应，本身说明国际上普遍希望与中国各层级对话交流。此外国际规则也迫切需要中国参与和认可。由于中国不是经济合作与发展组织成员国，经合组织的人工智能原则只能通过 G20 平台得到中国的认可。因此在当前形势下，需要积极建立国际对接渠道，有序推动如中国 – 经济合作与发展组织、中国 – 欧盟、中国 – 东盟乃至中美谈判。建议在外交场合推荐更多中国科学家和企业家参与相关专家委员会，确保规则制定的科学性。多头并举履行大国担当，确保我国在国际合约制定的前期讨论中不缺席。

除了参与交流，中国也需要更积极地引领国际对话。引领国际人工智能对话的基础是研究能力。2019 年有许多人工智能治理研究机构成立，如多伦多大学的施瓦茨·赖斯曼技术与社会研究所、华盛顿特区的安全与新兴技术中心，与早前成立的牛津大学人工智能伦理研究所和牛津大学互联网学院发布的新兴技术治理项目、人类未来研究所、未来生命研究所等，组成了全球最为重要的人工智能治理专业研究机构。我国近些年也有不少人工智能治理的研究机构成立，为对话交流奠定了基础。未来中国需要基于人工智能的研发和产业实践以及人工智能治理的研究基础，更加积极主动设置国际议题，引领国际对话。

3.3 创新治理路线，鼓励企业进行微观治理

从《2019 观察》来看，企业既是人工智能治理的受体，也可以是主体。企业可以率先开发相关的技术，来预防或者检验人工智能的恶意应用。以欧洲为主的人工智能伦理治理路线，采用自上而下的模式，先制定人工智能伦理“宪章”，再制定具体规则。但是从中国的技术产业发展现状和技术治理历史背景来看，我们或许可以采用其他路线。中国的企业不仅是被治理的对象，同时也是把握用户

需求的前哨。目前腾讯、旷视等中国公司虽然也提出了自己的人工智能伦理原则，但是从公开的信息源来看并没有具体的技术和制度方案，如果仅仅有原则而无可执行的方案，则会有“道德清洗”的嫌疑。如果企业能够开发出新的技术屏蔽人工智能假新闻、假视频，确保人工智能负责任发展，那将不仅仅是引领人工智能治理的走向，同时也有可能引领人工智能的产业方向。鉴于中国人工智能企业的快速发展，如果从市场中及时发现治理问题，然后积极探索技术和制度上的解决方案，有可能为人工智能产业的健康发展探索出合适的路线。■

参考文献：

[1] 上海市科学学研究所 . 我所发布《全球人工智能治理年度观察 2019》（英文版）[EB/OL]. [2020-12-28]. http://www.siss.sh.cn/c/2020-04-30/611360.shtml.

[2] AI Governance in 2019. A year in review: observations from 50 global experts[EB/OL]. [2020-12-28]. https://www.aigovernancereview.com/.

[3] Dafoe A. AI governance: a research agenda. Governance of AI Program[R/OL]. [2021-02-22]. https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf.

[4] Montreal AI Ethics Institute. State of AI Ethics[R]. Montreal: Montreal AI Ethics Institute, 2020.

[5] Brundage M, Avin S, Wang J, et al. Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims[R]. arXiv preprint arXiv: 2004.07213.

[6] Floridi L. Establishing the rules for building trustworthy AI[J]. Nature Machine Intelligence, 2019, 1(6): 261-262.

[7] 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（4）：6-8.

[8] Delcker J. Europe’s silver bullet in global AI battle: Ethics[EB/OL]. (2019-03-17) [2021-02-22]. https://www.politico.eu/article/europe-silver-bullet-global-ai-battle-ethics/.

[9] Stop Secret Surveillance Ordinance. Be it ordained by the people of the city and county of San Francisco[EB/OL]. (2019-05-07) [2021-02-22]. https://www.eff.org/deeplinks/2019/05/san-francisco-stop-secret-spy-tech-and-face-surveillance.

The Main Contents and Enlightenment of AI Governance in 2019: A Year in Review

LI Hui, Brian TSE, WANG Ying-chun

(Shanghai Institute for Science of Science, Shanghai　200235)

Abstract: Based on AI governance in 2019: a year in review, observations from 50 global experts, this paper analyzes the opinions of 44 representative institutions in the report on the progress of artificial intelligence governance in 2019, and points out that the current global multi forces are contributing to the “global artificial intelligence governance system”. This paper also studies the development trend and characteristics of the current global AI governance system incubation period, and finds some trends and characteristics, such as the formation of national community of AI governance, the transformation of governance rules from “soft law” to “hard law”, and enterprises becoming the leading force of micro governance. Finally, based on these trends and characteristics and the current situation of China’s AI governance, this paper puts forward some suggestions for China to participate in and further promote the construction of global AI governance system.

Keywords: artificial intelligence; global governance; governance system; international consensus