

人工智能对国际关系的影响初析

傅莹

【内容提要】 本文重点分析人工智能如何从国际格局和国际规范两个方面影响国际秩序的变迁。在国际格局方面,人工智能有可能在经济和军事领域影响国家间的力量对比,非国家行为体的能力也将前所未有的地扩大,围绕科技的国际竞争将更加激烈。在国际规范方面,人工智能有可能改变战争的形式和原则,并对现行的国际法律和伦理道德造成冲击。文章认为,人工智能技术带来的安全和治理挑战是需要全人类共同面对的问题,各国应从构建人类命运共同体的视角看问题,从共同安全的理念出发,讨论未来人工智能的国际规范。文章提出基于福祉、安全、共享、和平、法治、合作的六项原则。

【关键词】 人工智能 国际格局 国际规范 共同规范

【作者简介】 傅莹,清华大学兼职教授、国际关系研究院荣誉院长、国际战略与安全研究中心主任。

电子邮箱: fying@tsinghua.edu.cn

清华大学国际战略与安全研究中心于2018年启动的人工智能与安全项目小组,重点探讨人工智能技术的发展对国际关系的影响以及构建共同准则的可能性。本文是基于对项目组相关信息的汇总而对人工智能与国际关系的相互影响的初步研究。

1950年,英国科学家图灵提出了人工智能的概念。^①1956年,首届人工智能研讨会在美国新罕布什尔州达特茅斯举行,人工智能作为一门科学正式为科学界所承认。1997年,IBM计算机程序“深蓝”赢得与世界国际象棋

^① 1950年,图灵发表论文《计算机器与智能》(*Computing Machinery and Intelligence*),提出“图灵测试”的概念,即如果人类测试者在向测试对象询问各种问题后,依然不能分辨测试对象是人还是机器,那么就可以认为机器是具有智能的。

特级大师卡斯帕罗夫的六盘大战。进入 21 世纪的第二个十年,人工智能技术的研究和开发加快了步伐。2014 年,谷歌 AlphaGo 战胜世界围棋冠军李世石。60 年后的今天,人工智能在人类越来越多的生产和生活领域得到广泛应用,在部分专门领域接近甚至超过人脑的表现。作为一种有潜力改造人类社会面貌的泛在性技术,人工智能在科技、产业、军事、社会、伦理等领域被广泛讨论。

那么,人工智能是否会对国际关系产生影响?会产生什么样的影响?本文试图对这些问题作一些探讨。需要说明的是,人工智能技术本身存在复杂性、难以说明性和不确定性,笔者非人工智能技术专家,本文只是根据已经发生的人工智能事件或学术界普遍认可的发展趋势,来分析人工智能对国际关系产生的影响,并试图探讨构建共同准则的必要性和可能性。

诚然,对于科学技术会如何影响现代国际关系有不少过高的预期。例如,阿尔温·托夫勒在 1980 年出版的《第三次浪潮》一书中预测,未来世界将充斥核武器风险,濒临经济和生态崩溃,现存政治制度将迅速过时,世界将面临严重危机。这类预测往往高估了科技给人类带来的困难,却低估了人类解决困难的意愿和能力。冷战后,在全球化的大背景之下,多边主义逐渐成为国际共识,国际防止核武器扩散体系的有效性,应对气候变化的全球运动和各国由此而不断强化的应对环境变化的合作,以及世界范围内的和平运动的发展,都显示了人类在维护和平和应对挑战上的理念共识和负责任态度。科技造成的问题,可以通过科技本身的不断再进化而得到解决,人类也需要通过伦理道德和法律来构建严密的防范体系。实际上,每一次科技革命都加速了全球化的进程,促使一系列全球性问题被纳入国际政治的议事日程中来,世界也因此变得更加透明和更加融为一体。

一、我们讨论的是什么?

在进入正式讨论之前,需要明确几个问题。

第一个问题,我们讨论的是什么人工智能?是狭义的、能够模拟人的个别智能行为的人工智能,比如识别、学习、推理、判断?还是通用人工智能,拥有与人类大脑相似的自主意识和自主创新能力?是弱人工智能,为解决

特定、具体任务而存在,只擅长语音识别、图像识别、翻译某些特定素材的人工智能,例如谷歌的 AlphaGo、科大讯飞的智能翻译器?还是强人工智能,能够思考、计划、解决问题,能够进行抽象思维、理解复杂理念、快速学习、从经验中学习等人类级别的人工智能,例如电影《人工智能》中的小男孩大卫、《机械姬》中的艾娃?抑或是未来的超级人工智能,即跨过了“奇点”^①,计算和思维能力远超人脑,“在几乎所有领域都比最聪明的人类大脑聪明很多,包括科学创新、通识和社交技能”(牛津哲学家 Nick Bostrom 对超级智能的定义)的所谓人工智能合成(Synthetics)?

我们讨论人工智能对国际关系乃至国际格局的影响,只能限定在已知、基于大数据和深度学习技术,以算力、算法和数据为三大要素的人工智能技术及其应用上。我们还无法讨论那些尚未获得突破的、拥有全大脑仿真技术的未来人工智能技术。目前应是依据已经存在和能大致预见的人工智能技术及其发展趋势,来探讨其已经和可能对国际关系产生的影响。

第二个问题,人工智能能否影响国际关系进而影响到国际秩序?目前看,答案是肯定的。历史上,技术革新和传播曾经无数次革命性地改变了一国或多国命运,进而改变了地区格局甚至世界形势。麦克尼尔父子在《麦克尼尔全球史:从史前到 21 世纪的人类网络》一书中,生动地描写了技术革命对军事、政府组织方式、信仰,进而对国家间权力转移和地区格局演变的决定性影响。书中谈到,公元前 1700 年左右,战车革命改变了美索不达米亚、埃及、印度、中国黄河地区等区域的权力格局,例如,雅利安人入主印度北部和商朝的兴起。公元前 1200 年后,铁冶炼技术的出现和传播,使装备着相对价格低廉的铁甲胄和兵器的普通步兵,有条件将精英驾驭的战车掀翻在地,更加廉价的武器装备、更大规模的军队、官僚统治的巩固等条件的组合,使亚述、波斯等农业帝国的崛起成为可能。公元前 7 世纪,马弓手的数量和技术又一次打破欧亚大陆的军事政治平衡,大草原上的游牧民族再度取得对

^① 1990 年,美国未来学家雷·库兹韦尔在《奇点临近》《人工智能的未来》两本书中,用“奇点”作为隐喻,描述人工智能的能力超越人类的某个时空阶段。当人工智能跨越“奇点”后,一切我们习以为常的传统、认识、理念、常识将不复存在,技术的加速发展会导致“失控效应”,人工智能将超越人类智能的潜力和控制,迅速改变人类文明。

农耕民族的优势。^①

再举一例,核技术的出现改变了现代世界的政治面貌,进一步稳固了二战结束时形成的大国权力格局,五核国的身份使得美国、苏联(解体后是俄罗斯)、英国、法国、中国保持了联合国安理会常任理事国地位,而且产生了为和平目的利用核能、有核国家承担不扩散核武器义务、允许无核国家获取和平核能技术等一系列国际规范,催生了核战争等于人类毁灭、核扩散不义且非法等国际价值观,形成了《不扩散核武器条约》《全面禁止核试验条约》、联合国核裁军谈判机制、全球核安全峰会、东南亚无核区等一系列国际制度安排。

人工智能被认为具有像核能一样的军民两用性和改变国际实力对比的颠覆性。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了《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技术。^②因此,将人工智能纳入能够影响国际关系的讨论范畴是合理的。

人工智能甚至可能影响国际秩序的变迁。北京大学王缉思教授认为,国际秩序包含两项基本内容,一是主要国家和国家集团的权力结构和实力对比,二是处理国家间关系应遵循的规范。^③清华大学阎学通教授认为,国际秩序是“国际体系中国家依据国际规范采取非暴力方式解决冲突的状态”,其构成要素为国际主流价值观、国际规范和国际制度安排。^④他还认为,导致国际秩序变化的原因是国际格局的变化,但国际格局却不是国际秩序的构成要素;建立国际新秩序的性质是国际权力的再分配,即国际制度再安排的核心内容。^⑤两位学者在分析国际秩序时都强调了国际格局、国际规

① 约翰·R·麦克尼尔、威廉·H·麦克尼尔:《麦克尼尔全球史:从史前到21世纪的人类网络》,王晋新等译,北京大学出版社,2017年,第72—79页。

② Greg Allen and Taniel Chan, *Artificial Intelligence and National Security* (Cambridge: Belfer Center for Science and International Affairs, Harvard Kennedy School, 2017), p. 1.

③ 王缉思:《世界政治的终极目标》,北京:中信出版集团,2018年,第30页。

④ 阎学通:《无序体系中的国际秩序》,《国际政治科学》2016年第1期,第13—14页。

⑤ 同上,第10页,第15页。

范两大因素。人工智能有可能通过改变国际行为体的力量对比和相互关系,冲击现有国际规范并催生新的国际规范,从而从国际格局和国际规范两个方面影响国际秩序的变迁。

二、人工智能将如何影响国际格局?

首先,人工智能将从经济上影响国家间的力量对比,甚至引发新一轮大国兴衰。

保罗·肯尼迪在《大国的兴衰》一书中指出,从长远看,在每个大国经济的兴衰与其作为一个世界性大国的兴衰之间有一种显而易见的联系。^① 2017年6月,普华永道发布的《抓住机遇——2017夏季达沃斯论坛报告》预测,到2030年,人工智能对世界经济的贡献将达到15.7万亿美元,中国与北美有望成为最大受益者,总获益相当于10.7万亿美元。2018年9月,麦肯锡全球研究所发布的《前沿笔记:用模型分析人工智能对世界经济的影响》报告认为,人工智能将显著提高全球整体生产力。去除竞争影响和转型成本因素,到2030年,人工智能可能为全球额外贡献13万亿美元的GDP增长,平均每年推动GDP增长约1.2%。^② 这堪比或大于历史上其他几种通用技术(比如19世纪的蒸汽机、20世纪的工业制造和21世纪的信息技术)所带来的变革性影响。报告还指出,占据人工智能领导地位国家和地区(以发达经济体为主)可以在目前基础上获得20%~25%的经济增长,而新兴经济体可能只有这一比例的一半。“人工智能鸿沟”可能会导致“数字鸿沟”的进一步加深。^③ 人工智能可能改变全球产业链。以工业机器人、智能制造等为代表的“新工业化”将吸引制造业“回流”发达经济体,冲击发展中国家人

① 保罗·肯尼迪:《大国的兴衰》,北京:中信出版集团,2013年,前言第12页。

② McKinsey Global Institute, “Notes from the Frontier: Modeling the Impact of AI on the World Economy,” *Discussion Paper*, September 2018, p. 1, <https://www.mckinsey.com/~media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx>, 访问时间:2019年2月15日。

③ Ibid., p. 34.

力资源等比较优势,使许多发展中国家提前“去工业化”或永久性失去工业化的机会,被锁定在资源供应国的位置上。人工智能的开发和应用需要大量资金,科技含量高,且有可能导致就业结构变革,使得高重复性、低技术含量的工作逐渐消失。

此外,麦肯锡在2017年的另外一份报告中,根据对46个国家和800种职业进行的研究作出预测,到2030年,全球将有多达8亿人会失去工作,取而代之的是自动化机器人;届时,全球多达1/5的劳动力将受到影响。全球将发生类似在20世纪初的大规模岗位转变,当时全球大部分岗位从农业转为工业。^①简言之,就是用资本和技术替代劳动力。同时,人工智能技术的广泛应用也将增加对这方面的专业人才的需求。

研究显示,有3种类型的国家最有可能从人工智能技术的发展中受益。第一类是有人工智能先发优势的国家,比如美国和中国,都被看好。第二类是资本和技术密集、同时人口较少或处于下行趋势的国家,比如日本、韩国、新加坡,既有发展人工智能的资本、技术条件,又能借助人工智能的发展弥补人口总量不足或呈下降趋势、人口结构老龄化等劣势。第三类是拥有更多科学家、数学家、工程师的国家,或重视科学、技术、工程、数学(STEM)相关专业教育的国家。

其次,人工智能将从军事上改变国家间的力量对比。

军事智能化的鼓吹者认为,人工智能将颠覆战争形态和战争样式。机械化战争是以物释能,靠的是石油和钢铁;信息化战争是以网络聚能,靠的是信息和链接。而根据目前的预期,一旦战争进入智能化时代,将是以智驭能,打的是机器人和自动化战争。

可以预见的是,在智能化条件下,战斗人员、作战概念、制胜机理等战争要素都将发生改变。在传统的战争中,即便存在敌对双方在武器装备、训练水平上的差距,劣势一方凭借有利的天时地利、超人一等的计谋、先进的战术等,尚可一战。比如,在伊拉克和阿富汗战争期间,简易爆炸装置就让美

^① Rich Miller, “Robots Are Coming for Jobs of as Many as 800 Million Worldwide,” *Bloomberg*, Nov 29, 2017, <https://www.bloomberg.com/news/articles/2017-11-29/robots-are-coming-for-jobs-of-as-many-as-800-million-worldwide>, 访问时间:2019年2月16日。

军吃尽了苦头。而在智能化战争条件下,一方在人工智能上的技术优势会迅速形成战场上的压倒性优势,使劣势一方无法形成有效的观察—判断—决策—行动循环,始终处于被动挨打的状态。布鲁金斯学会在《人工智能改变世界》报告中提出了“极速战”(hyperwar)的概念,即战争是一个与时间赛跑的过程,决策最快和执行最快的一方通常会占上风。在人工智能辅助下的指挥和控制系统,其决策速度会大大超越传统战争模式,加上能够自主决定发射致命武器的自动武器系统,将大大加速战争进程,以至于需要创造一个新术语“极速战”来描述这种战争模式。^①关于后者,2018年4月26日“大西洋”网站刊发的《当排爆机器人变成武器》一文中,就用大量分析说明,军事机器人可以显著降低路边炸弹的威胁。^②

人工智能还将引发军事装备的革命性变化,无人自动性致命武器的集群式作战,可能成为未来战争的主角和主要作战方式。设想中的空中无人机“蜂群”、水下无人潜艇“狼群”、地面机器人、无人坦克“蚁群”等一旦出现,将重构“消耗战”“人海战术”的价值,使得舰母、F-35战斗机等复杂而昂贵的大型武器平台,从战争成本和作战效能的综合角度看变得不那么具有优势。可以想象一下,当单架成本上亿美元的F-35战斗机对战单价在几千美元的武装无人机集群时,无异于“大炮打蚊子”。

需要说明的是,人工智能对军事的影响尚存在很大不确定性,这种影响可以有多大、如何影响,都并不确定。在2018年7月清华大学第七届世界和平论坛“人工智能与安全”分论坛的讨论中,有专家提出,虽然目前大致可以判断出机器学习、工业机器人、材料科学等技术的未来发展趋势,但这些技术结合起来给未来战争造成的具体影响尚无法准确预估。20世纪前30年,当时的德、英、法、意等欧陆军事强国都研发出了坦克、飞机、无线电通信技

^① Darrell M. West and John R. Allen, “How Artificial Intelligence Is Transforming the World,” *Brookings Institution*, April 2018, <https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/>, 访问时间:2019年2月16日。

^② Caroline Lester, “What Happens When Your Bomb-Defusing Robot Becomes A Weapon,” *The Atlantic*, April 2018, <https://www.theatlantic.com/technology/archive/2018/04/what-happens-when-your-bomb-defusing-robot-becomes-a-weapon/558758/>, 访问时间:2019年2月16日。

术等。但是,只有当德国在二战中遂行了“闪电战”后,世界才发现这些新技术作为一个整体竟会给战争带来如此颠覆性的改变。现在,无论“算法战”还是“蜂群”战术,战略界的热议都还是从单项技术出发分析其对作战的影响。如果不能从整体上认识人工智能技术的军事运用,设想出来的应对措施有可能成为昂贵而无用的新“马其诺防线”。

再次,按照目前的研究成果看,人工智能对国际格局的影响可以概括为三个方面。

一是可以显著增加全球财富。但是财富在国家间的分配将更加不平衡,人工智能技术的引领国受益多,落后国受益少,人工智能将进一步拉大国家间经济总量和质量差距。新增财富在地区间的分配也将是不平衡的,东亚地区有可能成为人工智能技术的最大受益者,美国有条件巩固其世界经济中心的地位,而东亚的地位也会进一步上升。与此同时,财富在同一国不同地区、不同行业、不同群体中的分配可能进一步失衡,相关国家国内政治平衡也将受到不同程度的冲击。二是在资本、技术、劳动力三要素中,人工智能有利于具有前两项要素比较优势的国家,不利于劳动力要素优势的国家。三是当前人工智能技术的核心要素是算法、算力和数据,哪个国家拥有的计算资源越多、研发的算法越先进、掌握的数据越多,就越有可能获得经济优势。未来可能出现“算法博弈”“算法霸权”,数据将成为国家战略资源。四是由于人工智能技术迭代速度快,在“马太效应”作用下,具有先发优势的国家将强者恒强,后发国家越来越难以追赶,形成国际社会的“阶层固化”。由于当前人工智能技术的发展建立在计算机技术进步基础之上,并与数字经济发展存在紧密关联,因此数字经济和技术领先的国家将在人工智能国际竞赛中占得先机。

最后,人工智能带来的经济和军事变化,在影响国家间力量变化的同时,还可能意味着国际趋势的变化。

一是北升南降还是南升北降?近年随着金融危机影响消退,发达经济体逐步复苏,而金砖国家相继出现困难,发展中国家追赶步伐放慢,南升北降的趋势开始减缓。如果不论其他因素,单就人工智能技术的发展而言,很可能加剧这一势头。一方面,发达经济体有条件维持甚至扩大对发展中国家的整体优势。另一方面,发展中国家之间的实力、利益分化也可能进一步加剧。

二是多极化还是两极化？阎学通教授认为国际格局的趋势是中美两极化，而不是多极化。^① 人工智能技术的发展为他的观点提供了一定的支持。国际上很多人工智能排名都不约而同地将美国和中国列为第一梯队。美国作为人工智能的发源地，在创新精神、基础理论、核心算法、高端芯片、从业人员规模和素质等方面，明显地领先全球。而中国在人脸识别、语音识别等领域走在前面，同时拥有海量数据和应用场景等其他国家难以比拟的天然优势。中国拥有全球最多的互联网用户、最活跃的数据生产主体。2018年8月20日，中国互联网络信息中心发布的第42次《中国互联网络发展状况统计报告》显示，截止2018年6月，中国网民规模达到8.02亿人，每年产生的数据约占全球13%。^② 中国拥有更多的人工智能相关专业大学生，人才供给上更具潜力。根据腾讯研究院发布的《中美两国人工智能产业发展全面解读》报告，截至2017年6月，全球人工智能企业总数为2542家，其中美国有1078家，占42%，从业人员约为7.8万人；中国有592家，占23%，从业人员为3.9万人；其余872家企业分布在瑞典、新加坡、日本、英国、澳大利亚、以色列、印度等国家。^③ 人工智能技术竞赛目前主要是巨头企业之间的角力，而主要的互联网巨头也都在将自己的资源转向人工智能。在世界互联网巨头企业中，美国有苹果、谷歌、微软、亚马逊、脸书5家，中国有百度、阿里巴巴、腾讯3家。美中作为人工智能的世界前两个强国，同时也是全球综合国力的前两强，与其他国家的差距可能会进一步拉大。

三是非国家行为体的权力扩大。人工智能将赋予巨头企业前所未有的权力。2016年发生的“剑桥分析”公司通过脸书影响美国大选的事件，显示了这些企业可以有多大的政治和社会影响力。未来，一些传统的政府职能，如信息统计和收集、公众信息发布、社会福利发放还有公共政策讨论、评估

① Yan Xuetong, "Unipolar or Multipolar? A Bipolar World Is more Likely," *China-US Focus*, Vol. 6, April 2015, pp. 12-15.

② 中国互联网络信息中心：《中国互联网络发展状况统计报告》，2018年8月，第20页，http://www.cac.gov.cn/2018-08/20/c_1123296882.htm，访问时间：2019年2月18日。

③ 腾讯研究院：《中美两国人工智能产业发展全面解读》，2017年7月26日，第3页，http://www.tisi.org/Public/Uploads/file/20170802/20170802172414_51007.pdf，访问时间：2019年2月18日。

与反馈,以及涉及军事和外交方面的舆论活动等,都可能被人工智能企业取代。在信息即权力的时代,企业行为对国内政治、进而对国际政治的影响力还将不断增大。美国斯坦福大学和美国 Infinite 初创公司联合研发了一种基于人工智能处理芯片的自主网络攻击系统。该系统能够自主学习网络环境并自行生成特定恶意代码,实现对指定网络的攻击、信息窃取等操作。通过人工智能自主寻找网络漏洞的方式,将使网络作战行动更加高效,攻击手段更加隐蔽和智能。传统的防护方式是基于病毒库和行为识别,将无法应对灵活多变的人工智能病毒生成系统,其恶意代码的生成、执行、感染具有更强的隐蔽性,这将使网络安全环境面临更大的挑战。^①

更令人担心的是,人工智能也有可能为恐怖分子、网络黑客、罪犯等提供新的犯罪手段。例如,恐怖分子可以远程操纵无人机或无人驾驶汽车,对目标实施暗杀和破坏活动,黑客可以利用大数据武器对关键基础设施进行更加复杂和自动化的大规模攻击。经常被提到的案例包括:俄罗斯在叙利亚的赫梅米姆和塔尔图斯军事基地遭受武装分子无人机攻击,委内瑞拉总统马杜罗在演讲时遭受无人机袭击,“想哭”病毒的爆发,等等。加拿大学者阿查亚提出了“复合世界”概念。^②毫无疑问,人工智能将增加世界的复合化,包括政治权力的去中心化、更多元的行为主体的出现,乃至政治和意识形态的多样性,因此产生更加包容和更多层次的治理需求。

四是加剧科技竞争。人工智能技术的发展及其应用,进一步巩固了科技在国际竞争中的重要地位。2016年5月,美国国家科技委员会国土与国家安全分委会发布的《21世纪国家安全科学、技术与创新战略》提到,科技已经成为一个国家国际竞争力最关键的成分。没有科技创新就没有国家安全。美国能够在国际竞争中保持优势,很大程度上有赖于其科技创新力量。国内许多学者认为,中美贸易战的实质是科技战,中美博弈已经率先在科技领域展开了。2018年,美国先后对两家中国公司禁售芯片,更有甚者,公开阻止盟友国家采用中国华为公司的5G技术。同年11月19日,美国商务部

^① 学术 Plus:《美军人工智能武器化大盘点》,搜狐军事,2019年1月7日,http://www.sohu.com/a/287109476_100044418,访问时间:2019年2月19日。

^② 阿米塔亚·阿查亚:《中国与自由主义国际秩序的危机》,《全球秩序》2018年第1期,第81页。

工业安全署列出了拟议管制的14个“具有代表性的新兴技术”清单,其中对人工智能和机器学习的技术分类和列管最为详尽。^①可以预见,在人工智能时代,围绕科技的国际竞争将更加激烈,竞争的结果也将在很大程度上影响国际格局的走向。

三、人工智能将如何影响国际规范?

人工智能可能对现行国际规范带来一系列冲击。

第一,人工智能有可能改变战争的内涵和形式。美国学者斯蒂芬·平克指出了人类相互间使用暴力减少的历史大趋势,认为“今天我们也许处于人类有史以来最和平的时代”^②。王缉思对当今世界战争明显减少的现象给出了5个方面的解释,包括:大国发动战争可能付出的成本和代价大大高于可能的收益,通过非战争手段也可以获得过去需要通过战争才能获得的收益,国家间相互了解的增加以及危机预防和管控机制的普遍建立,国际军控机制的建立,和平作为一种国际价值观的深入人心。^③而人工智能的发展有可能使国家和非国家行为体在是否使用军事手段解决矛盾的选择中,顾虑下降。至少从理论推演的结果看,人工智能能够从决策者、军队、民意3个层面减少对战争的制约,降低战争门槛。无人武器使得战争行为者的人员伤亡预期可以降到近乎“零”。人工智能技术的快速迭代和新材料的广泛运用将极大地压缩战争的物理成本,让战争的经济适用性上升,甚至变得有利可图。这使得决策者发动战争的诱惑增大、顾虑减少。“超视距”作战的可能性和机器人战士的出现,将进一步从心理上解放士兵。未来战争中,士兵或

^① 美国商务部“具有代表性的新兴技术”清单在人工智能和机器学习项下列出的技术包括:神经网络和深度学习(例如脑模拟、时间序列预测、分类)、进化和遗传计算(例如遗传算法、遗传编程)、强化学习、计算机视觉(例如物体识别、图像理解)、专家系统(例如决策支持系统、教学系统)、语音和音频处理(例如语音识别和制作)、自然语言处理(例如机器翻译)、规划(例如调度、博弈)、音频和视频处理技术(例如语音克隆、deepfakes)、AI云技术、AI芯片组。

^② 斯蒂芬·平克:《人性中的善良天使:暴力为什么会减少》,北京:中信出版集团,2015年,前言第1页。

^③ 王缉思:《世界政治的终极目标》,北京:中信出版集团,2018年,第55—59页。

许不必上战场,作战任务可以通过事先设计好的算法,由“人工智能将军”指挥无人自动武器完成。“极速战”可以极大压缩战争时间,在出现民意的反弹之前战争就可以结束了。这些可能改变人们对战争的立场,让暴力回潮。

2018年10月,美国陆军协会陆战研究所发布的报告《影响力机器——让自动化信息作战成为战略制胜机制》称,在人工智能的帮助下,利用算法生成内容,实施个性化的目标锁定和采用密集的信息传播组合,可以生成“影响力机器”,实施信息作战,由此能产生指数级的影响效应。该报告认为,“影响力机器”信息作战在战略层面上的影响力远胜于人工智能技术在其他领域的应用。因为它可以在机器学习的辅助下,对情感、偏见和价值观等指标进行筛选,并锁定那些心理最易受到影响的目标受众,然后将定制的“精神弹药”快速密集的“射向”目标群体,达到影响其心理、操纵其认知的目的。^①

第二,人工智能可能冲击全球战略稳定。约瑟夫·奈曾经谈到,核武器的出现使军事力量作为维护安全的手段走到了极限,甚至走向其反面。人们认识到,在核战争中没有赢家,核武器可能使全人类面临灭亡的危险。^②而人工智能将挑战经典的“相互确保摧毁”理论。兰德公司在2018年发布的《人工智能对核战争风险的影响》报告中分析认为,到2040年,人工智能技术的进步使报复性核反击力量成为目标并被摧毁的可能性大大增加,从而削弱“相互确保摧毁”的基础,打破核战略平衡。即使各国无意发起先发制人的攻击,也会倾向于追求先发制人的能力,以此作为与对手讨价还价的手段,而这无疑将破坏战略稳定。^③

第三,人工智能的自身特点让建立相关国际军控和防扩散机制变得困

① 学术 Plus:《美军人工智能武器化大盘点》,搜狐军事,2019年1月7日,http://www.sohu.com/a/287109476_100044418,访问时间:2019年2月19日。

② 约瑟夫·奈:《理解国际冲突:理论与历史》,上海人民出版社,第201—207页。

③ Edward Geist and Andrew J. Lohn, “How Might Artificial Intelligence Affect the Risk of Nuclear War?” *Rand Corporation*, p. 8. <https://www.rand.org/pubs/perspectives/PE296.html>, 访问时间:2019年2月19日。

难。在2018年7月清华大学第七届世界和平论坛“人工智能与安全”分论坛的讨论中,《智能时代的战略竞争》报告的作者之一格里戈利·艾伦(Gregory C. Allen)说,人工智能也有军民两用性,但其军事应用不同于核,更像电。如果可以认为核是一种黑白技术,只存在“有”或“没有”两种状态,一国要么是有核国,要么是无核国。而电则是一种光谱技术,就像红橙黄绿蓝靛紫七色光有不同的波长和媒介反射率一样,不同的国家都可以用电,只是应用方式和程度会有很大差异。人工智能在军事上的应用更像电,无法禁止哪一个国家使用人工智能。就像美苏冷战期间的军备竞赛一样,未来很可能出现人工智能强国之间的算法竞赛。问题是,当年美苏通过军控谈判签署了一系列核导军控协议,确定了基本规则,未来的人工智能强国间能否本着同样的精神达成算法控制协议呢?艾伦认为,从目前大国关系态势看,达成这样的共识几无可能;然而,考虑到未来无序发展的风险,大国又必须认真考虑就此进行共同探讨的必要性。

第四,人工智能给国际法带来了一系列的问题。在应用人工智能武器的条件下,国际人道法和战争法的有关原则是否能够继续适用?比如,区别对待军民目标的“区分原则”、禁止过分攻击的“比例原则”、非军事手段无法达成目标的“军事必要原则”,对作战手段的限制等。有没有必要针对人工智能武器制定专门的规则?在智能化战争条件下,如何区分战斗人员和非战斗人员?战争机器人是否享有人道待遇?人工智能武器是否对其造成的损害有承担责任的能力?如果它不具备承担责任的能力,那责任主体应该是武器的制造者还是使用者?当人工智能武器侵犯国家的主权原则时,它的行为能否触发国家责任?

第五,人工智能可能冲击国际关系民主化。人工智能的发展有可能固化国际权力结构,深化小国、弱国对强国、大国的科技、经济和安全的依赖。当“赢者通吃”的人工智能技术竞争和商业竞争规律复制到国际关系上时,必然会冲击大小、强弱国家间的主权平等关系。当主要大国之间的竞争加剧时,结盟关系是否会再度成为弱国的必然选择呢?而结盟本身就是一种盟主和盟友之间的不平等关系。

第六,人工智能对全球治理提出了新的课题。人工智能的发展对于解

决当今世界面临的三大困境(老龄化、数字化和气候变化)都具有意义。^①而更大的挑战也许是人工智能带来的对于“人”的价值的进一步思考。^②一些国家和城市也开始探索试行“普遍个人收入”制度(UPI, universal personal income)。^③但这必定是一个需要汇聚全球智慧与力量共同去思考和实践的问题。

在规则与制度层面上,以数据为例,在近代人类历史上,国际社会先后就自然人的跨境移动、船只和飞行器等交通工具的跨境移动、资本和商品的跨境移动等,形成了普遍认同的规则和制度安排。当前,数据作为一种资源的重要性与日俱增,随之而来的问题是:数据跨境转移应遵循什么规则、在什么样的制度安排下进行?数据的产生方、使用方、输出方、接收方等利益相关各方分别享有什么权利、承担什么义务?数据本土储存和跨境转移之间是什么关系?国家与数据企业之间是什么关系?这些都将是未来全球治理的难题。2019年G20的东道主日本已经提议将全球数据治理列入G20的议程。

① Kathryn Reilly, “UK’s Ageing Crisis Has Become the ‘New’ Climate Change: Can Digital Innovation Save the Care System from Collapsing?” *MedTech Engine*, <https://medtechengine.com/article/uks-ageing-crisis-has-become-the-new-climate-change-can-digital-innovation-save-the-care-system-from-collapsing/>, 访问时间:2019年2月20日; McKinsey Global Institute, “Digitization, AI, and the Future of Work: Imperatives for Europe,” *Briefing Note*, September 2017, p. 1, [https://www.mckinsey.com/~media/McKinsey/Featured Insights/Europe/Ten imperatives for Europe in the age of AI and automation/Digitization-AI-and-the-future-of-work.ashx](https://www.mckinsey.com/~media/McKinsey/Featured%20Insights/Europe/Ten%20imperatives%20for%20Europe%20in%20the%20age%20of%20AI%20and%20automation/Digitization-AI-and-the-future-of-work.ashx), 访问时间:2019年2月20日。

② Pricewaterhouse Coopers, “Human Value in the Digital Age,” December 2018, <https://www.pwc.nl/nl/assets/documents/pwc-human-value-in-the-digital-age.pdf>, 访问时间:2019年2月20日。

③ Annie Nova, “More Americans Now Support A Universal Basic Income,” *CNBC*, Feb 26, 2018, <https://www.cnn.com/2018/02/26/roughly-half-of-americans-now-support-universal-basic-income.html>, 访问时间:2019年2月20日; Peter S. Goodman, “Finland Has Second Thoughts about Giving Free Money to Jobless People,” *New York Times*, Apr 24, 2018, <https://www.nytimes.com/2017/07/20/opinion/finland-universal-basic-income.html>, 访问时间:2019年2月20日。

四、我们该如何选择？

目前,即便走在人工智能技术最前列的科研人员也承认,人类距离制造出达到人类智慧的机器人,还有比较长的距离。目前,人工智能可能带来的安全威胁更多是人类本身利用人工智能去威胁人类。因此,人类如何进行自我约束是人工智能技术进步过程中最为紧迫的道德问题。

2015年10月14日,一个名为“智能平方”的组织在纽约邀请几位国际问题专家进行了一场辩论,题目是:中美是长期敌人吗?。芝加哥大学政治学教授米尔斯海默不出所料地讲起了“大国政治的悲剧”。作为反方,澳大利亚前总理陆克文有一句话令人记忆犹新。他说,外交的挑战就是要确保阻止战争的发生,相信我们能够做到这一点。^① 陆克文的逻辑也适用于人工智能。人工智能是一种技术,可以为善,也可以作恶。尽管对善、恶的判断和选择是人类一个古老和难有完全一致看法的哲学命题,但是在关系到人类根本生存这样的大问题上,拥有理性和现代文明的人类应该可以决定向善还是向恶。

2018年6月,笔者应邀参访科大讯飞北京公司,其间讨论到未来是否会出现人工翻译与机器翻译之间的竞争时,公司负责人表示,发展人工智能翻译技术的目的不是取代人,而是助力翻译工作更加准确和轻松。这给笔者以启发。毕竟,就像毛泽东同志所说的“决定战争胜败的是人民,而不是一两件新式武器”^②。人类开发人工智能技术的目的是服务于人类,是善用还是恶意应用最终取决于掌握技术的人。

当前国际上关于秩序的讨论十分热烈。其中,中美两国学界最为关心的,也为世界其他国家所普遍关注的是:美国与中国是否会展开“新冷战”,从而导致冷战时的平行秩序再度出现? 2018年3月,笔者与来北京参加中国发展高层论坛的美欧人士交谈时,观察到他们对中美关系的前景比较悲

^① IQ2US Debates, “China and the U. S. Are Long-term Enemies,” October 14, 2015, <http://intelligencesquaredus.org/debates/past-debates/item/1403-china-and-the-u-s-are-long-term-enemies>, 访问时间:2019年2月22日。

^② 《毛泽东选集(第四卷)》,北京:人民出版社,1991年,第1195页。

观,认为中国的快速增长对美国来说是结构性威胁。当笔者问道:中美两国除了走向对抗,有没有别的选择?英国《金融时报》副主编马丁·沃尔夫回答说,你们没有什么可选择的,中国不会停止增长,美国不会停止担忧。(You have no choice, you can't stop growing and they can't stop worrying.)除非出现一种情况,例如发生来自火星的入侵,人类面临共同的敌人,届时美中才有可能真正团结起来。

如果用零和博弈和追求绝对安全的眼光看待世界,那毫无疑问,人工智能会像20世纪40、50年代的原子弹、卫星一样,成为大国竞争的新焦点,并成为将世界划分为两种或多种平行秩序的助推动力。但是如果我们采用人类命运共同体的视角,本着共同安全的理念看问题,那么就不难认识到,人工智能技术带来的安全和治理挑战是需要全人类共同面对的问题。如此,我们就不难本着平等协商的精神,共同探讨各利益攸关方都能够接受的规范。那么,人工智能是否会成为那个将中国、美国、俄罗斯以及世界其他国家团结在一起的“火星入侵”式的挑战呢?

技术专家们在大声疾呼。在2018年的乌镇第五届世界互联网大会上举办的“网络空间的中美关系”分论坛上,原美国联邦通信委员会专家戴夫·法伯呼吁,应尽早就人工智能制定国际准则,避免重演核武器技术产生时因国际共识和自律的迟到而导致的悲剧及其影响至今的严重后果。

现实情况是,中美两国在科学技术研究领域的交往与合作是相当深入的。根据科睿唯安提供的人工智能领域科技文献数据,从2013年到2017年,全球参与评估的167个国家/地区在WOS学科“计算机、人工智能”发文排名中,中国大陆位居第一,发文量达到59573篇,占25.02%,其次是美国,发文量为32527篇,占13.66%。其中,中美两国国际合作论文数量增长最快,例如,过去5年中国大陆开展国际合作最多的对象是美国,合作论文量是4307篇,同样,美国开展国际合作最多的国家也是中国大陆,远远多于与其他国家的合作论文量。^①

国际上围绕相关问题的探讨和研究已经在步步深入。例如,联合国裁

^① 科睿唯安信息服务:《人工智能领域科技文献中高产国家/地区的竞争力分析》,2018年12月,第13页。

军委致命性自主武器系统问题政府专家组的讨论就涉及要注意致命性自主武器系统可能对国际安全产生的影响。例如,致命性自主武器系统技术方面的军备竞赛,扩大发达国家与发展中国家之间的技术差距,以及可能降低使用武力的门槛等;就国家对其管辖下的致命性自主武器系统负有法律责任达成非正式普遍谅解;探讨参照《特定常规武器公约》议定书的形式,制定一项具有法律约束力的文件,预防性地禁止致命性自主武器系统。

习近平主席于2018年9月在致2018世界人工智能大会的贺信中阐明了中国的立场和态度。他指出,新一代人工智能正在全球范围内蓬勃兴起,为经济社会发展注入了新动能,正在深刻改变人们的生产生活方式。把握好这一发展机遇,处理好人工智能在法律、安全、就业、道德伦理和政府治理等方面提出的新课题,需要各国深化合作、共同探讨。中国愿在人工智能领域与各国共推发展、共护安全、共享成果。^①

基于清华大学人工智能与安全项目小组的研究结果,我们提出了关于人工智能的六点原则:一是福祉原则。人工智能的发展应服务于人类共同福祉和利益,其设计与应用须遵循人类社会基本伦理道德,符合人类的尊严和权利。二是安全原则。人工智能不得伤害人类,要保证人工智能系统的安全性、可适用性与可控性,保护个人隐私,防止数据泄露与滥用。保证人工智能算法的可追溯性与透明性,防止算法歧视。三是共享原则。人工智能创造的经济繁荣应服务于全体人类。构建合理机制,使更多人受益于人工智能技术的发展、享受便利,避免数字鸿沟的出现。四是和平原则。人工智能技术须用于和平目的,致力于提升透明度和建立信任措施,倡导和平利用人工智能,防止开展致命性自主武器军备竞赛。五是法治原则。人工智能技术的运用,应符合《联合国宪章》的宗旨以及各国主权平等、和平解决争端、禁止使用武力、不干涉内政等现代国际法基本原则。六是合作原则。世界各国应促进人工智能的技术交流和人才交流,在开放的环境下推动和规范技术的提升。这些原则可以作为讨论和制定人工智能国际规则的基础。

^① 新华社:《习近平致2018世界人工智能大会的贺信》,2018年9月17日,http://www.xinhuanet.com/politics/leaders/2018-09/17/c_1123441849.htm,访问时间:2019年2月22日。

虽然这些原则尚显空泛和抽象,如何细化和如何与各国专家深入探讨,寻找最大公约数,还是一个需要花时间去努力争取的目标。值得注意的是,许多从事高端科研的青年专家已经提出为人类的共同利益而主动自律的问题,不少企业界人士表示要在人工智能的技术研究和产品制造中自觉强调道德和道义的因素,不会从事有损人类福祉的研究和制造工作。

笔者希望国际上关于这些问题的探讨能不断深入,也期待中国的智库和科学技术界能在这个方向上为全球的共同努力贡献力量。