

人工智能时代的国际关系： 走向变革且不平等的世界

封 帅

摘 要 深度学习算法的突破使人工智能技术进入了新一轮的高速发展周期，依托于算法和算力的全面提升，人工智能技术已经迈过产业化的门槛，成为推动新一轮产业革命的关键力量。人工智能技术的快速发展将对国际关系领域产生深远的影响。首先，人工智能技术已经直接卷入战略决策和军事领域的国际互动中，将进一步放大各行为体之间的力量差距，打破原本稳定的均势结构。其次，人工智能技术所带来的自动化生产会改变全球社会经济生产模式，推动国际行为体内部的权力分配结构变革，最终对国际体系产生重大的系统性影响。最后，人工智能技术的发展还会影响新时代的思想观念，新的社会思潮将在不断的碰撞与辩论过程中逐渐成型。人工智能时代的到来将使人类进入一个变革且不平等的世界，中国也将迎来新的机遇与挑战。

关键词 人工智能 深度学习 战略博弈 军事系统 自动化生产 国际关系变革

* 封帅，上海国际问题研究院国际战略研究所助理研究员（上海 200233）。

** 本文系上海市发展与改革委员会“全球多语种信息监测与决策分析平台项目”（项目编号：HYJY_20160816）的阶段性成果。感谢陈志瑞教授及匿名评审专家对于本文提出的重要修改意见。本文在写作过程中，曾与上海国际问题研究院院长助理叶青、复旦大学博士后刘树才以及周亦奇、王玉柱等师友就相关问题进行过深入的讨论。此外，在相关学术会议上，华东师范大学余南平教授、苏州大学陈文亮教授提出的很多观点对笔者也很有启发。谨此一并致谢。

我们从来没有像今天这样接近那个曾经无数次在幻想中出现的未来。随着计算机硬件性能的全方位提升和互联网时代海量数据的积累，经历了多次波折的人工智能技术在 21 世纪第二个十年进入了新一轮高速发展期。本轮人工智能技术的高速发展以人工神经网络的研究为基础，有赖于深度学习等算法方面的重要突破。^① 依托于算法和算力的空前提升，人工智能在诸多垂直领域都取得了兼具商业开发潜力和社会影响力的卓越成果，技术的发展首次跨越了产业化的门槛，被世界各国学者和企业界视为推动新一轮科技革命的关键力量。

作为一项划时代的科技成果，人工智能与人类社会的历次科技革命都有特质上的差异。旧的技术革新无论形态如何，其性质仍是人类改造世界的工具和手段，而人工智能则能够通过通过对大数据的分析和学习，理解人类的内在需求，作为创造性的伙伴直接参与人类改造世界的活动，并表现出与人类理性思维方式完全不同的思考逻辑，这在一定程度上改变了人类与技术的深层次联系，也将对国际关系领域产生深远的影响。

一方面，人工智能在与国际关系密切相关的特定领域内的应用，将赋予技术拥有者以额外的优势，从而改变诸多领域的权力结构。另一方面，人工智能技术在诸多垂直领域的渗透，在创造更多社会财富的同时，也潜移默化地重塑人类社会经济活动的方式，重构人类的社会组织形态，推动国际行为体的内部变迁，进而引发国际体系的深刻变革。随着人类社会逐渐步入人工智能时代，国际体系的主体、结构、运行规则等关键要素都将随之发生巨变，传统的国际关系理论也将面临严峻的考验。因此，充分了解人工智能领域的前沿发展态势，分析未来人工智能时代的基本特征，研讨人工智能技术的高速进步对国际关系领域的潜在影响，对于我们理解未来国际体系可能的演变前景，以及国家间互动模式的变迁，具有重要的意义。

^① 关于深度学习算法的相关问题，参见：Geoffrey E. Hinton, Simon Osindero and Yee-Whye Teh, “A Fast Learning Algorithm for Deep Belief Nets”, *Neural Computation*, Vol. 18, No. 7, July 2006, pp. 1527-1554; Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, 2015; Yoshua Bengio, Pascal Lamblin, Dan Popovici, Hugo Larochelle and Greedy LayerWise, “Training of Deep Networks”, in J. Platt, et al., eds., *Advances in Neural Information Processing Systems* (NIPS 2006), MIT Press, 2007, pp. 153-160; Marc Aurelio Ranzato, Christopher Poultney, Sumit Chopra and Yann LeCun, “Efficient Learning of Sparse Representations with an Energy-Based Model”, in J. Platt, et al., eds., *Advances in Neural Information Processing Systems* (NIPS 2006), pp. 1137-1144 [美] 伊恩·古德费洛、[加] 约书亚·本吉奥、[美] 亚伦·库维尔 《深度学习》，赵申剑等译，人民邮电出版社，2017 年。

20世纪80年代末至90年代初,部分美国和西方学者就初步探索了人工智能技术对于国际关系领域的影响,但囿于当时人工智能的技术水平,这种讨论并未在学科内部获得广泛关注,相关研究主要集中于如何利用人工智能开展数据库建设或进行辅助决策,而对于国际关系的本原性问题涉及较少。近两年来,随着人工智能技术的飞速进步,相关研究开始呈现出明显的增长态势。^①我国社会科学界对于人工智能问题关注较晚,早期对于人工智能的思考多出现于哲学和经济学领域。近两年来,部分学者在推动探索国际关系领域的预测问题时开始涉及相关问题,但直接以人工智能与国际关系问题为目标的研究成果尚未出现。^②

基于此,本文将以前即将到来的弱人工智能时代的基本特征为出发点,尝试分析当前人工智能技术对于国际关系领域产生的直接影响、系统性的间接影响以及对于国际关系理念的长期影响,努力勾勒在弱人工智能时代国际体系要素变革的宏观图景。除此之外,笔者也希望对该问题的初步研究能够抛砖引玉,唤起更多国际关系研究者对这一重要问题的关注,在人工智能的发展史上留下中国国际关系学者的思想印记。

^① 关于人工智能与国际关系的研究参见: Stephen J. Cimbala, ed., *Artificial Intelligence and National Security*, Lexington Books, 1987; Allan M. Din, ed., *Arms and Artificial Intelligence: Weapon and Arms Control Applications of Advanced Computing*, Oxford University Press, 1988; Valerie M. Hudson, ed., *Artificial Intelligence and International Politics*, Westview Press, 1991; Ennals Richard, *Artificial Intelligence and Human Institutions*, Springer-Verlag, 1991; George F. Luger, *Artificial Intelligence, Structures and Strategies for Complex Problem Solving*, Addison-Wesley, 2005; Erik Brynjolfsson and Andrew Mcfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, Norton, 2014; Martin Ford, *Rise of the Robots: Technology and the Threat of a Jobless Future*, Basic Books, 2015; Jerry Kaplan, *Humans Need Not Apply: A Guide to Wealth and Work in the Age of Artificial Intelligence*, Yale University Press, 2015; Kareem Ayoub and Kenneth Payne, "Strategy in the Age of Artificial Intelligence", *The Journal of Strategic Studies*, Vol. 39, No. 5-6, 2016, pp. 793-819; Vincent Boulanin and Maaike Verbruggen, *Mapping the Development of Autonomy in Weapon Systems*, Stockholm International Peace Research Institute, 2017; Greg Allen and Taniel Chan, "Artificial Intelligence and National Security", Intelligence Advanced Research Projects Activity (IARPA), Belfer Center, Harvard Kennedy School, July 2017.

^② 迄今在中国国际关系学界与人工智能问题直接相关的研究成果尚未出现,不过部分成果对人工智能的相关问题已经有所涉猎:董青岭《机器学习与冲突预测——国际关系研究的一个跨学科视角》,《世界经济与政治》,2017年第7期,第100—117页;庞珣《国际关系的定量方法:定义、规则与操作》,《世界经济与政治》,2014年第1期,第5—25页;庞珣《定量预测的风险来源与处理方法——以“高烈度政治动荡”预测研究项目的再分析为例》,《国际政治科学》,2017年第1期,第1—32页。

一、弱人工智能时代的技术影响力扩散模式

现代意义上的人工智能研究最早可以追溯到 20 世纪 40 年代。1950 年，英国著名科学家图灵在《计算机与智能》^①一文中首次提出了“机器能够思考吗？”这一具有划时代意义的理论问题，并同时提出了测试机器是否拥有智能的方法。

1956 年，在达特茅斯夏季学术研讨会上，研究者们接受了计算机专家约翰·麦卡锡提出的人工智能（AI）概念，并将其作为这门新兴学科의 正式标签。^②然而，在随后几十年的时间里，人工智能的发展却并非一帆风顺，虽然在不同时期出现了“专家系统”、“深蓝”等多项具有标志性意义的成果，但由于客观条件的限制，人工智能技术始终无法有效解决人们的现实需求，技术发展无法在产业层面落地，对于现实社会的影响非常有限。

真正的突破出现在 2009—2010 年前后，硬件设备的进步使新一代计算机在运算速度和信息处理能力方面得到大幅提升。互联网产业的发展改变了人类的生活方式，使得网络成为人们获取日常生活数据最集中、最便捷的渠道。移动互联时代的到来，则使来自网络搜索、电子商务、社交媒体、科学研究等不同领域的海量数据迅速累积，为人工智能的飞跃提供了充足的养分。在已具备强大计算能力与大数据环境的情况下，早在 60 年代就已经被提出的多层神经网络工具焕发出巨大的生命力，重新成为技术发展的主流路径。^③本轮人工智能发展热潮以“深度学习”为核心，该算法以建立套嵌式的多层次模式识别系统组成的“神经”架构为基础，通过组合低层特征形成更加抽象的高层属性、类别或特征，借以发现数据的分布特点。^④深度学习的出现带来了人工智能算法的跨越，改变了传统符号主义学派以计算机模拟

^① Alan M. Turing, “Computer Machinery and Intelligence”, *Mind*, Vol. 59, No. 236, October 1950, pp. 433-460.

^② John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude E. Shannon, “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence”, *AI Magazine*, Vol. 27, No. 4, 2006, pp. 12-14. 关于达特茅斯会议的倡议，可见 http://www.universelle-automation.de/1956_Dartmouth.pdf。

^③ 关于神经网络研究的历史，可参见〔美〕伊恩·古德费洛、〔美〕约书亚·本吉奥、〔美〕亚伦·库维尔《深度学习》，第 8—12 页。

^④ Geoffrey E. Hinton, Simon Osindero and Yee-Whye Teh, “A Fast Learning Algorithm for Deep Belief Nets”.

人类认知系统推进人工智能的艰难尝试，让人工智能拥有了从海量且复杂的信息源中提取、识别和构建体系的能力，在那些任务目标明确且相关数据丰富的领域，深度学习算法能够让机器学习新的技能，制定有效策略，从而在短时间内提出超过人类学习能力的问题解决方案。^①

随着深度学习神经网络成为主流发展方向，人工智能技术很快在现实场景中得以应用。在很短的时间内，人工智能技术就在图像识别、语音识别、机器翻译、自动驾驶甚至棋类竞赛等复杂的应用场景中获得了飞速的进步，基本达到了满足人类日常需求的标准，具备了商业投资的价值，并很快成为资本市场追捧的新热点。迈过产业化的门槛意味着人工智能技术真正走出了实验室，能够对社会生产和人类生活产生直接影响。更重要的是，本轮人工智能技术的进步证明，机器学习算法能够在具有很多限定性条件的领域比人类做得更好，并且能够通过自我学习不断进步。这一结论事实上打开了人类对于人工智能发展的未来想象空间，人类社会已经能够清晰地听到未来社会匆匆而至的脚步声。

本轮在深度学习引领下的人工智能技术发展，将使人类社会逐渐步入被称为“弱人工智能”（Artificial Narrow Intelligence, ANI）时代的发展阶段，在部分文献中也将其称为领域性人工智能（Narrow AI）阶段、应用性人工智能（Applied AI）阶段或模块化人工智能（Modular AI）阶段。本轮人工智能技术的重要成果由三个基本要素相互融合而成，即硬件设备的快速进步、大数据的持续积累、深度学习算法的不断突破。理论上说，在硬件设备持续进步的条件下，在所有拥有大数据的垂直领域，人工智能技术都能够通过算法的更新为具有明确目标指向的问题提供解决方案，并形成有价值的应用性成果。

然而，多层神经网络的特性决定了深度学习模型是通过“训练”与“学习”来解决问题的，模型设计者本身并不能准确地知道该模型在解决问题时发现了哪些特殊规律，系统对于人类而言是一个完全的“黑箱”。正如杰瑞·卡普兰所说，“（机器学习系统）发展出自己的直觉，然后用直觉来行

^① 从理论流派来说，深度学习应该属于人工智能领域的“联结主义”（Connectionism）学派，部分人工神经网络的学者也以这个概念称呼自己，但这种归类方式本身也存在一定争议。关于人工智能领域的不同理论路径，可参见 Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*。

动。”^① 这也就意味着能够跨领域解决问题的通用人工智能（General AI）几乎不可能在现有的理论框架中出现，据美国国家科学技术委员会预测，通用人工智能的出现还需要几十年的时间，^② 因此，人类社会将在未来相当长的一段时间内处于弱人工智能时代。我们也需要根据弱人工智能时代的影响力扩散方式来分析其对于国际关系的影响。

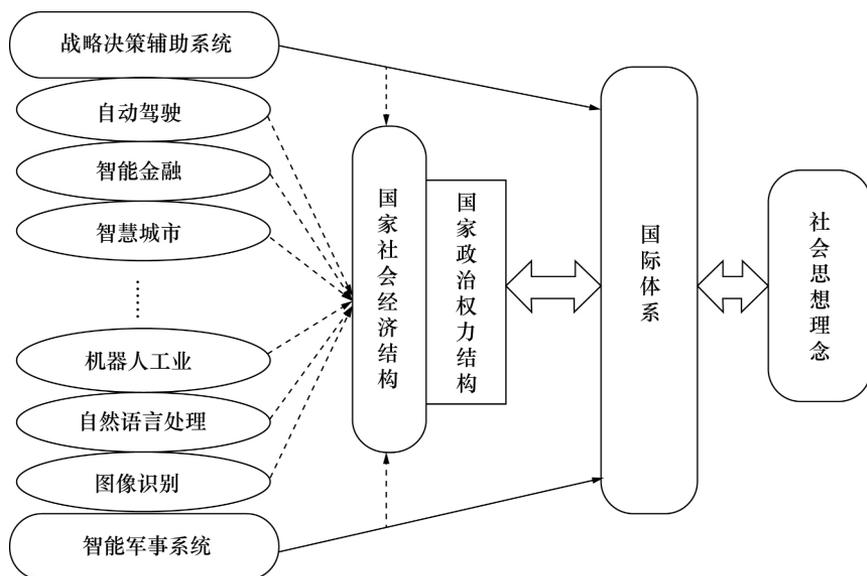


图-1 弱人工智能时代技术影响力的扩散模式

如图-1 所示，从当前的发展态势来看，在弱人工智能时代，深度学习所带动的领域性人工智能技术的持续进步会在三个层次上对国际体系产生影响。

其一，从领域层面上看，部分人工智能技术可以直接应用于特定国际关系议题领域，从而改变现有各行为体在该领域的力量对比，为国际体系增加新的不稳定因素。目前在人工智能技术的推进过程中，新技术的应用主要集中在经济收益较高的民用领域，直接针对国际政治活动的研究仍然处于起步阶段，但在战略决策辅助系统和智能军事系统两个领域，人工智能已经能够

^① Jerry Kaplan, *Humans Need Not Apply: A Guide to Wealth and Work in the Age of Artificial Intelligence*, p. 31.

^② Executive Office of the President, National Science and Technology Council, Committee of Technology, "Preparing for the Future of Artificial Intelligence", October 2016, p. 7, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

成为国际政治活动参与者的重要辅助性工具，而随着技术水平的不断提升，在该领域的技术代差将对国家力量对比产生不可逆的影响。

其二，从制度层面上看，人工智能技术的领域性扩散将对现代社会的经济发展模式产生重要的影响，重塑世界经济体系中资本与劳动力之间的关系，进而推动国家治理模式与权力分配方式的演变，最终影响国际体系的稳定，但所有领域性的人工智能技术进步都指向一个共同的目标，即实现“以机器替代人类劳动”，这也就意味着在经济体系中资本力量的快速扩张和劳动力要素的持续削弱，在新的技术条件下，全球经济体系将出现结构性的调整。社会经济条件的改变会直接影响原有政治体系的稳定，而传统的科层制政府体制需要进行重大调整才能有效适应新的经济基础，全球经济与治理体系的变化最终将引发现有国际体系的深刻变化。

其三，从思想层面上看，跨时代技术变革的影响会逐步扩展到整个社会生活，并最终反映在政治与社会思想的变迁进程中。而这些思想理念也将反过来影响国际关系和国内治理。面对近年来人工智能技术的快速发展，人类社会正进入一个思想领域相对困惑而迷茫的阶段。一方面，世人对人工智能技术进步的成果孜孜以求。另一方面，却又无时无刻不在担忧人工智能对人类文明未来可能形成的反噬。阿隆·马斯克与史蒂芬·霍金所发出的对于人工智能风险的警告就是对这种担忧最直接的反应。^① 人类社会的价值与思想理念将在弱人工智能时代面临重构的压力，由于不同领域和学科对于这一问题的认知存在分歧，各种理念的碰撞将长期存在。重塑人类社会的思想理念与价值标准，也将成为社会科学研究介入人工智能议题的重要路径。

依托上述分析框架，我们可以更加清晰地讨论弱人工智能时代技术进步对于国际关系各要素所产生的复杂影响，并尝试勾勒未来国际关系的宏观图景。

二、人工智能的领域性介入：战略博弈与军事系统

当前，人工智能技术与国际关系直接相关的应用主要集中在战略决策系统和军事系统两个领域，这是由国际环境与弱人工智能时代的技术特点共同

^① Elon Musk, "Artificial Intelligence Is Our Biggest Existential Threat", *The Guardian*, October 27, 2014; Rory Cellan-Jones, "Stephen Hawking Warns Artificial Intelligence Could End Mankind", *BBC News*, December 2, 2014, <http://www.bbc.co.uk/news/technology-30290540>.

决定的。一方面，战略决策与军事安全关系国家的根本利益，所有国家都会努力提升自己在这两个领域的能力，以便在国际关系活动中取得优势地位，因而各国对于新技术的研发与应用始终保持积极的态度；另一方面，这两个领域议题目标相对明确、敌我关系相对清晰、数据积累较为丰富，能够满足深度学习算法发挥作用的所有客观条件。因而顺理成章，这两个领域成为连接人工智能与国际关系的桥梁，成为弱人工智能时代新技术介入国际关系的优先领域。

（一）人工智能与战略博弈模式的变革

利用人工智能技术推动建立和完善战略决策辅助系统，是弱人工智能时代技术变革对国际关系领域的最直接影响之一。战略博弈领域天然适合人工智能技术发挥其比较优势。人工智能技术从诞生之日起，就一直将确定规则及目标的博弈行为作为重要的理论突破口，以阿尔法狗（AlphaGo）战胜围棋冠军柯洁为标志，人工智能已经成功证明了自己在所有具备“完全信息博弈”特征的游戏项目中拥有难以匹敌的压倒性优势。而在其他如德州扑克等“不完全信息博弈”的竞技比赛中，人工智能在策略制定方面的优势甚至更加明显。^①

其实，从数学意义上看，国际行为体间的战略互动过程同样是一种非常典型的“不完全信息博弈”行为。参与者通过判断博弈对手的能力、意图、利益和决心，并结合对特定外部环境的分析，制定出最为有利的博弈策略并加以实施。^② 国际关系领域的战略博弈涉及范围更广，内容更加复杂，各项要素相互联系所形成的系统效应，实际上已经远远超出人类思维所能够分析和掌控的范畴。在传统意义上，国家参与战略博弈的过程更多依赖政治家的直觉与判断，而深度学习算法的飞速进步，则使得人工智能决策辅助系统逐步显示出其在该领域的明显优势和发展潜力。只要技术进步的大趋势不改变，人工智能将注定改变现有的国际战略博弈过程，国际行为体之间的战略关系也将发生重大变革。^③

从当前深度学习算法的特点来分析，人工智能战略决策辅助系统将提升

^① 关于不完全信息博弈问题的基本模式，参见 John C. Harsanyi, “Games with Incomplete Information Played by ‘Bayesian’ Players, Part I: The Basic Model”, *Management Science*, Vol. 14, No. 3, November 1967, pp. 159-182.

^② 关于战略博弈的分析过程，参见唐世平《一个新的国际关系归因理论：不确定性的维度及其认知挑战》，《国际安全研究》，2014年第2期，第3—41页。

^③ Kareem Ayoub and Kenneth Payne, “Strategy in the Age of Artificial Intelligence”, pp. 793-819.

战略决策的科学性、有效性与可靠性，进而推动战略博弈模式的深度变革。

首先，人工智能系统能够提供更加精确的风险评估和预警，使战略决策从一种事实上的主观判断转变为精确化的选择过程，从而提升战略决策的科学性。战略博弈的前提和基础是对于前景和风险的准确预测与判断，但任何人类的决策行为本身都是根据极为有限的情报信息进行的具有较强主观色彩的猜测。同时，人类决策过程始终处于信息超载的状态下，决策者只能将部分信息作为战略判断的主要依据。根据认知心理学的观点，人类更容易吸纳那些符合自身偏好的信息，决策者也会因为固有的知识体系的影响，为自己建立认知捷径。这种普遍存在的状态会造成两种结果：有时决策者会由于自我认知的偏见而对各个领域的状态做出过于乐观的判断，有时又会因为信息触及其认知捷径的盲区，而夸大低概率事件出现的可能性。总之，传统的风险评估和预警具有明显的模糊性与偶然性特征。^①

但当深度学习技术介入战略博弈过程之后，博弈活动本身将被人工智能系统抽象为一个贝叶斯网络模型。^②在大数据环境下，深度学习算法可以通过监督学习或非监督学习的方式，分类归纳相关信息，并设立风险参数，通过逐步建立基于人工智能技术的风险预警系统，较为准确地预测博弈过程中的系统性风险。^③在人工智能技术介入战略博弈进程之后，已经可以通过技术手段以概率的方式精确表达风险严重程度与应对策略的有效性。在深度学习技术介入战略博弈之后，决策过程可以从主观猜测变为从不同概率的前景中挑选策略。风险和策略的量化表达最大限度减少了战略博弈过程中的不确定性，有助于实现决策的科学化。^④

其次，深度学习算法能够以更快的速度提供更多不同于人类常规思维方式的战略选项，并且随着博弈过程的持续，进一步根据对方策略的基本倾向完善本方策略，提升战略决策的有效性。战略决策本身是在复杂的环境下对所能获取的不完整信息进行分析、最终做出判断的过程。但在深度学习算

① Daniel P. Kahneman, *Thinking, Fast and Slow*, Allen Lane, 2011, pp. 89-96.

② Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, Chapter 6.

③ 参见 Jochen Kruppa, "Risk Estimation and Risk Prediction Using Machine-Learning Methods", *Human Genetics*, Vol. 131, No. 10, October 2012, pp. 1639-1654; 董青岭 《机器学习与冲突预测——国际关系研究的一个跨学科视角》，第100—117页。

④ Jorge Galindo and Pablo Tamayo, "Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications", *Computational Economic*, Vol. 15, No. 1-2, April 2000, pp. 107-143.

法介入之后，参与博弈的决策过程就真正实现了对于“大数据”的深度利用。人工智能系统可以利用科技力量，以人类无法比拟的工作效率阅读和分析数据，并通过对机器的“学习”和“训练”，在短时间内掌握人类分析人员需要用较长时间训练才能获得的情报分析能力。

更为重要的是，人工智能的决策模式与人类思维模式完全不同，它是在给定目标体系的情况下，利用多层次神经网络系统构建新的博弈策略。深度学习算法的特点决定了机器智能并不是模仿人类的决策过程，而是在充分理解目标的基础上主动构建新的策略。这些策略通常与人类的思维方式差异较大，但在实现博弈目标方面往往更加有效。深度学习的特点使得人类往往无法理解人工智能系统的思考方式，只能通过策略的执行结果进行推断，但该系统在现有以博弈模型为基础的竞技类游戏中都能够轻易地取得明显优势，并且找到人类思维的很多盲点加以有效利用。^①

深度学习算法的另一个重要优势是可以从错误中学习，并且在反复互动过程中更加准确地掌握博弈对手的策略特征。随着积累的数据量不断扩大，博弈对手的形象会逐渐清晰起来，其主要思考方式及策略特点也会被逐渐纳入机器的思考范围内。在较长的周期内，即便人类决策者最初能占据一定主动，但随着博弈进程的延续，也必然会因为策略被理解而逐渐丧失优势。^②

最后，在战略博弈进程中，人工智能系统能够最大限度地排除人为因素的干扰，提高战略决策的可靠性。在参与战略博弈的过程中，人类会受到特定文化和社会心理的影响，但人工智能是一个完全客观和中立的决策体系，代表着纯粹理性的态度。人工智能决策体系只会受到参数的影响，不会出现激情驱动决策的现象。它不会因为自己对于“荣誉”等主观因素的渴望或者对于某些未知因素的“恐惧”而影响自己的决定，甚至也不会考虑与战略目标相冲突的道德因素。当处于战略优势环境时，人工智能不会轻敌自满，做出错误决策。而处于劣势状态时，人工智能决策系统也不会变得更加冒险。即便是从战略博弈最为原始的方面来看，人类决策者在考虑战略决定时，会受到个人心理素质、身体状况、抗压能力甚至环境舒适度的影响。^③ 人类永远无法摆脱荷尔蒙或葡萄糖对于身体的干扰，但人工智能不会因为疲劳而做出任何改变，因此，无论从哪个角度来看，人工智能系统都是最为可靠的决

① Kareem Ayoub and Kenneth Payne, "Strategy in the Age of Artificial Intelligence", p. 808.

② *Ibid.*, p. 815.

③ Stephen P. Rosen, *War and Human Nature*, Princeton University Press, 2005, pp. 27-70.

策者，是理性选择的最完美诠释。

当然，在现有算法条件下，人工智能技术在战略博弈活动中也存在一定的局限性。一方面，在现实的战略博弈过程中，人类决策者往往会追求动态的目标，具体的战略目标体系会随着现实利益关系的变化而不断调整。事实上，在很多博弈过程中，博弈参与方所追求的目标往往是模糊的、变动的，且是无法衡量的。这些模糊状态通常被视为人类战略博弈中的精华内涵，而能否把握不同目标定位之间的微妙关系常常被认为是衡量决策者智慧的标尺，但深度学习算法的特点要求在博弈过程开始前必须提前设定战略目标，并且以此对系统展开“训练”，人工智能无法理解既有目标之外的“非意图性影响”，从而在现实政治活动中会遭遇很多意外挑战。^①

另一方面，深度学习算法需要依托大数据，但在弱人工智能时代，数据搜集过程本身需要人工参与，而人类自身的意图会潜移默化地影响数据搜集过程，最终形成的数据集会将数据搜集者固有的倾向与偏见保留下来，在这样的数据环境中，人工智能的比较优势往往会受到侵蚀，而来自数据搜集者的偏见最终也将体现在战略博弈的策略中，干扰决策的有效性。

总的来说，以人工智能技术为基础的决策辅助系统在国际战略博弈的进程中将发挥重要作用，技术的完善将使得国际行为体之间战略博弈能力的差距进一步扩大。如果战略博弈中的一方拥有人工智能技术，而另一方没有，那么这种战略博弈几乎就成为一场单向透明的对抗。缺少人工智能技术辅助的行为体将在风险判断、策略选择、决策确定与执行效率以及决策可靠性等多个方面处于绝对劣势，整个战略博弈结构将会完全失衡。

（二）人工智能技术在军事领域的扩散

军事领域是人工智能技术参与国际关系活动的另一个重要领域。早在人工智能技术诞生之初，美国军方就对人工智能技术在军事领域的应用给予充分的重视。^②事实上，人工智能技术与军事领域之间具有天然的联系，军事领域是引领人工智能技术突破的重要平台，很多重要人工智能的研发最初都是为了军事目的。而且由于军事活动的复杂性，目前几乎所有的深度学习算

^① Amos Tversky and Daniel Kahneman, *Judgment under Uncertainty: Heuristics and Bias*, The Hebrew University and Oregon Research Institute, 1973.

^② 参见 Allan M. Din, ed., *Arms and Artificial Intelligence: Weapon and Arms Control Applications of Advanced Computing*; Jeffrey L. Caton, "Autonomous Weapons Systems: A Brief Survey of Developmental, Operational, Legal and Ethical Issues", Strategic Studies Institute, U. S. Army War College, December 2015, <http://www.strategicstudiesinstitute.army.mil/pdffiles/PUB1309.pdf>.

法在垂直领域的突破都可以在军事领域得以应用。人工智能技术的进步将使武器系统、军事策略、军事组织甚至战争的意义发生深刻改变，人类社会也将在进入弱人工智能时代之后迎来一个完全不同的军事安全环境。^①

首先，在军事技术层面，人工智能技术的介入将推动智能化武器的广泛应用。如果说人类早期推进无人机或作战机器人研究的出发点是希望通过远程操控减少士兵伤亡，那么在深度学习算法出现之后，智能化武器的高速发展就开启了自主作战系统阶段。

早在2004年，美国研究团队利用类似深度学习的算法成功“训练”机器人系统在人类未干涉的情况下学会了驾驶直升机。^② 该项目标志着智能化武器系统建设成为可能。近年来，随着深度学习算法的进一步完善，原本以远程操控为特点的无人机、作战机器人等自动化武器系统都逐渐进入智能化开发阶段。利用深度学习算法，可以使智能化武器在虚拟环境中建立武器操控的基本能力，随后在现实环境中根据数据反馈不断提升战斗能力，学习执行各种战斗命令。智能化武器的规模越大，其在战斗中相互协调的优势就越容易发挥出来。通过共同的算法进行“训练”的大批量智能化武器，可以借助人工智能系统协调行动，优化作战策略，并且根据战场形势和作战目标进行灵活调整，最大限度地获得战场优势。^③ 智能化武器系统不仅能够使人与武器实现实质性分离，更将战争活动完全变成武器系统的任务，使实际伤亡率无限趋近于0，而且能够最大限度地提升武器的使用效率和不同武器之间的协作。更重要的是，智能化武器的使用将使“消灭敌人有生力量”等传统作战法则失去实际意义，而当人工智能武器系统与传统模式的军事力量在战场上相遇，战争的法律与伦理观念将受到极大冲击。

其次，在军事战略层面，人工智能技术的介入将使军事战略与策略水平达到前所未有的高度。除了武器系统之外，人工智能技术对于军事战略制定

^① Vincent Boulanin and Maaïke Verbruggen, *Mapping the Development of Autonomy in Weapon Systems*, pp. 16-17.

^② 该项目是由当时在斯坦福大学任教的吴恩达 (Andrew Ng) 领导的团队完成的，随后吴恩达先后在谷歌和百度公司任职，对推动全球人工智能技术的进步作出了重要贡献。参见 Andrew Y. Ng, et al., "Autonomous Inverted Helicopter Flight via Reinforcement Learning", in M. H. Ang and O. Khatib, eds., *Experimental Robotics IX*, STAR 21, March 2006, pp. 363-372.

^③ Vincent Boulanin and Maaïke Verbruggen, *Mapping the Development of Autonomy in Weapon Systems*, pp. 27-29; Peter Singer, *Wired for War: The Robotics Revolution and Conflict in the Twenty-first Century*, Penguin, 2009, pp. 124-125.

和军事指挥方面的改变更具革命性的意义。相较于复杂的人类社会，战争本身是相对简单且封闭的领域。深度学习算法已经能够较为完整地还原全部战场信息，模拟双方兵力部署和作战能力，完成相对精确的战场沙盘推演。^①在战争开始之前，通过无数次计算，人工智能系统能够完成对战场形态的模拟分析，并以概率形式对所有可能出现的前景进行量化，最终根据各种军事策略的量化概率选择作战方案，进行有效的兵力分配和战略部署。

在实际的战场环境中，人工智能具有更全面高效地搜集战场信息的能力。利用机器视觉可以提升自动武器系统对于战场情况的识别和分析能力，利用自然语言处理系统可以更有效地搜集和处理音频信号，自动驾驶技术可以使武器系统获得更高的灵活机动性和自主判断处理问题的能力。在这种情况下，战斗指挥员能够更全面和准确地掌握战场信息，获得更加精确的战术建议。随着人工智能技术的不断介入，真正的自主作战系统正不断走向成熟。

在军事指挥方面，人工智能拥有两个人类无法比拟的优势：其一，人工智能系统可以快速处理战场信息，具有人类所不具备的快速反应能力；其二，人工智能系统具有多线程处理能力，可以处理同时展开的多项军事行动，提出人类思维模式所无法理解的复杂策略。^②进入弱人工智能时代后，技术的影响将加剧常规军事力量对抗的不平衡状态，缺少人工智能技术辅助的武装力量将越来越难以通过战术与策略弥补战场上的劣势，常规对抗将不再是合理的战略选项，不对称战争将成为这两种力量对抗的主要方式。

最后，人工智能技术的深度介入将为全球核威慑体系的稳定带来新的变数。作为军事领域上一轮体系革命的主要推动力，核武器一直被视为当代全球军事力量平衡的基础，建立在“确保相互摧毁”原则基础上的核威慑体系在很大程度上仍然是现代国际体系安全稳定的基石。据此，战略界构建了经典的威慑理论，将国家的核力量规模、使用力量的意志与国家面对威胁的战略

^① Kevin Nelson, et al., "Evaluating Data Distribution and Drift Vulnerabilities of Machine Learning Algorithms in Secure and Adversarial Environments", *Proceeding of SPIE, Machine Intelligence and Bio-inspired Computation: Theory and Applications VIII*, 911904, May 22, 2014; Jason Scholz, et al., "Machine Learning for Adversarial Agent Microworlds", *MODSIM 2005 International Congress on Modeling and Simulation*, December 2005, pp. 2195-2120.

^② Li Yingjie, et al., "RTS Game Strategy Evaluation Using Extreme Learning Machine", *Soft Computing*, Vol. 16, No. 9, February 2012, pp. 1623-1637.

评估都作为要素纳入理论框架之中。^①然而，当人工智能技术介入核威慑体系的各个环节之后，原本稳定的体系会出现变动。

一方面，人工智能技术作为一种数据化工具，为网络战提供了前所未有的进攻性能力，对核武器使用的可靠性会产生直接的影响；另一方面，在大数据环境下，原本具有较大不确定性的国家意志与战略意图等主观因素，将能够在深度学习算法中以概率的形式获得明确展示，于是，原本处于“相互摧毁”状态的核威慑系统变得不再平衡。拥有人工智能技术的一方将具有清晰评估对方进行核反击的可能性及破坏性的能力，从而拥有更加灵活的战略选项，而技术相对落后的一方，其核反击能力将变得不再具有可信性。在弱人工智能时代，只有人工智能技术与核武器的结合才能形成有效的威慑系统。^②行为体之间的军事力量差距将因为人工智能技术水平再次扩大，国家的军事战略目标也将随之调整，传统的国际安全体系将变得不再稳定。

总的来说，人工智能技术对于军事领域的深度介入是自核武器发明以来最重要的军事技术变革之一。^③与以往的技术变革不同，人工智能技术对军事领域的改变是全方位的，从军事武器到战略设计、从全球军事力量平衡到军事伦理，都将不可避免地受到冲击。人工智能技术的进步将再次在短时间内放大军事主体间力量的差距，拥有方将具有全面超越传统军事力量的能力，使对方原本有效的伤害手段失效。新的不平衡状态可能会使“预防性战争”模式重新抬头，而寻找新的不对称作战方式则将成为中小国家军队或实力较弱的武装组织的主要发展方向。

在这种状况下，以人工智能技术为核心的新的军备竞赛将不可避免，大国必须追求人工智能技术并运用于军事领域，以维持其大国地位。中小国家则会千方百计寻求人工智能武器的扩散，争取获得新的制衡手段。当前，由于人工智能在军事领域的门槛较高，技术发展需要巨大的成本投入，短期内技术扩散的可能性较小，但随着技术的不断发展，其成本也将逐渐递减，技术扩散的风险随之提升，全球安全体系和伦理基础都将面临严峻的考验。

（三）走向更加不平等的世界

人类既有的历史经验告诉我们，任何科技革命的出现都会使率先掌握新

^① Henry A. Kissinger, *Nuclear Weapons and Foreign Policy*, Harper & Row, 1957; Thomas Schelling, *The Strategy of Conflict*, Harvard University Press, 1960; Thomas Schelling, *Arms and Influence*, Yale University Press, 1966; Robert Jervis, “Deterrence Theory Revisited”, *World Politics*, Vol. 131, No. 2, January 1979, pp. 289-324.

^② Greg Allen and Taniel Chan, “Artificial Intelligence and National Security”, pp. 18-20.

^③ *Ibid.*, pp. 15-18.

科技的国家与其他国家之间的力量差距进一步扩大。作为人类科技史上最有效的力量放大器，人工智能在战略博弈和军事等重要国际关系领域已经展现出明显超越人类的能力与持续发展的潜力，这种差距已经很难用数量堆砌或策略战术加以弥补，应用人工智能的行为主体几乎不可能在博弈过程中或军事行动中被尚未使用人工智能技术的对手击败，这将使得弱人工智能时代的国际主体间的力量鸿沟变得更加难以跨越。

同时，人工智能技术还会倒逼那些更早应用人工智能技术的国家在相关领域推动组织形式的改革。在人工智能技术获得广泛应用的背景下，科层制组织已经不再适应新的技术环境。传统上，军队组织体系和国家战略决策咨询体系都遵循严格的科层制原则，但人工智能技术以扁平化、网络化和高度协同化的方式推动体系运作，以等级和命令为核心的科层制体系完全无法适应这一新的形态。^① 随着人工智能技术在上述应用领域的不断扩展，传统的组织结构必将因新技术的采用而作出调整。当适应新技术的新体制推广开来，国际关系行为体之间将呈现出技术和制度上的系统性差距，发展中国家将面临更加严酷的国际竞争压力。

人工智能技术的发展将推动我们走向一个更加不平等的世界。快速扩大的力量差距使传统的国际关系主体间互动模式被迫发生重大改变，以主权国家平等为基石的国际社会的稳定结构也将面临严重的外部压力，构成弱人工智能时代国际关系领域变革的外部因素。

三、人工智能的系统性影响：新型社会经济结构与权力分配模式

在战略与军事领域的应用只是整个人工智能科技浪潮中的很小一部分，虽然具有改变国际行为体的力量对比的潜力，但毕竟还停留在从外部对现有国际体系构成冲击的阶段，如果人工智能技术的作用仅限于此，那么尚不足以对整个体系产生根本性的影响。然而，如图-1 所示，在人工智能技术影响力的扩散模式中，对国际关系最为深远的影响实际上是通过间接路径实现的。人工智能技术变革将从根源上对全球经济结构与政治治理模式产生系统

^① 关于人工智能与科层制治理结构之间的关系，参见贾开、蒋余浩《人工智能治理的三个基本问题：技术逻辑、风险挑战与公共政策选择》，《中国行政管理》，2017 年第 10 期，第 40—45 页。

性的影响，推动国际体系深层次的变革。这种冲击的直接着力点在于现代社会稳定的经济结构，经济结构的改变又将导致国际行为体内部权力分配方式的改变，最终由内而外推动体系的变革。

（一）弱人工智能时代的社会经济结构变迁

人工智能从诞生之日起，就注定是一项要对传统的社会经济体系发起全面挑战的技术。由具有与人类智慧相似的机器来代替人类从事生产工作，这本就是人工智能产业的核心逻辑。随着深度学习算法的广泛应用，人工智能技术正在越来越多的领域内向“机器替代人类劳动”这一基本目标迈进。自动化生产在各个领域的出现、发展和完善，构成了人工智能对世界经济体系最直接的影响。^①目前，在基础制造业领域，机器人生产已经逐渐获得行业的广泛认可，并成为制造业大国未来的主要发展战略。^②在金融服务领域，人工智能量化分析工具的普及也正逐渐压缩原本属于人类的高薪工作。随着无人驾驶、智能决策系统等新技术的逐渐成熟，仓储、交通运输、城市管理，乃至法律、教育和医疗等行业，都将被深深卷入由人工智能引发的技术变革浪潮中去。这是一场具有产业革命意义的技术变革，它将改变人类生产的基本形态，也将改变社会经济的基本结构。

毫无疑问，科技进步是社会发展的源动力，人工智能技术的广泛应用，最终将推动人类整体劳动生产率的提升，创造更加丰富的社会财富。其实，即便在人工智能技术尚未成熟的1993—2007年，17个工业国通过工业机器人所带动的自动化生产，已经使劳动生产率提升了约0.4%。^③在深度学习算法逐步完善之后，人工智能将使劳动生产率获得空前提升。据埃森哲的预测，到2035年，应用人工智能技术的企业利润将平均增长38%，在12个主要发达经济体的16个行业中实现14万亿美元的经济增长。^④但是，人工智

^① Executive Office of the President, “Artificial Intelligence, Automation and the Economy”, December 2016, <https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF>. 关于当前AI在自动化生产方面取代劳动力的情况，可参见Katja Grace, John Salvatier, Allan Dafoe, Baobao Zhang and Owain Evans, “When Will AI Exceed Human Performance? Evidence from AI Experts”, *Artificial Intelligence*, May 2017, <https://arxiv.org/pdf/1705.08807v2.pdf>.

^② 例如，近年来德国提出的“工业4.0”战略、中国提出的“中国制造2025”战略、美国提出的“工业互联网”战略，都是对这种趋势做出的反应。

^③ 参见Georg Graetz and Guy Michaels, “Robots at Work”, CEPR Discuss Paper No. DP10477, March 2015, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2575781.

^④ Mark Purdy and Paul Daugherty, “How AI Boosts Industry Profits and Innovation”, Accenture Research, 2017, http://img1.iyiou.com/Document/2017-07-11/Accenture_AI_Industry_Growth_Full_Report.pdf.

能技术具有革命性意义的一面在于，它是以颠覆原有产业结构和生产方式作为获得新的经济增长的基本路径的，因此，随着人工智能技术在社会生产活动中的全面推进，世界将不得不经历一个漫长的阵痛阶段，两种在当前人类社会中难以容忍的负面现象将在过渡阶段广泛出现，并成为经济社会结构的常态。

首先是持续性的失业潮。人工智能技术的核心逻辑是以机器自动化生产取代人类劳动，以提高社会生产率，因此，持续而普遍的失业现象将成为弱人工智能时代的基本特征。根据美国国家科学技术委员会的预测，在未来10—20年时间内，9%—47%的现有工作岗位会受到威胁，平均每3个月就会有约6%的就业岗位消失。^①与传统上基于生产规模下行所导致的周期性失业不同，由人工智能所导致的失业现象从本质上说是一种结构性失业，资本以全新的方式和手段替代了对于劳动力的需要，结构性失业人口将不会因为经济周期的变化重新获得工作，因为他们之前所能够适应的岗位已经彻底消失。主要依赖重复性劳动的劳动密集型产业和依赖信息不对称而存在的部分服务行业的工作岗位，将首当其冲被人工智能所取代。而随着人工智能技术在各个垂直领域的不断推进，受到威胁的工作岗位将越来越多，实际的失业规模会越来越大，失业的持续时间也会越来越长。^②

其次是社会财富的严重两极分化。在弱人工智能时代，与持续失业潮相伴而生的另一重要社会现象是社会财富的两极分化。这种两极分化的现象会在不同层次上反复出现，并对传统的社会结构造成严重的撕裂。其一，个体的财富分配会趋向两极分化，作为资本挤压劳动力的重要进程，人工智能所带来的劳动生产率的提升并不会转化为工资收入的普遍增长，换言之，自动化生产所创造的经济繁荣并不会惠及更多劳动人口。^③新创造的社会财富将会不成比例地向资本一方倾斜，低收入与受教育程度较低的人群将在新一轮的社会资源分配中处于严重的不利地位。其二，行业的财富分配也将呈现明显的分化状态，人工智能技术的运用将使拥有互联网与大数据条件的相关行业获得天量的资本注入，吸引更多的人才投入其中。在未来5—10年，人工智能相关产业与传统产业之间将形成巨大的发展鸿沟，无论是市值、资本还

① Executive Office of the President, "Artificial Intelligence, Automation and the Economy", p. 2.

② 参见 Martin Ford, *Rise of the Robots: Technology and the Threat of a Jobless Future*. 该书中文版可见(英)马丁·福特《机器人时代:技术、工作与经济的未来》，王吉美、牛筱萌译，中信出版社，2015年，第239—244页。

③ Executive Office of the President, "Artificial Intelligence, Automation and the Economy", p. 2.

是财富创造方面，都将出现严重分化。^①其三，从全球范围来看，国家间的财富分化将进一步加剧。人工智能技术的进步需要大量的前期投入，特别是在数据搜集和计算机技术方面的技术积累，对于人工智能产业的发展至关重要，但各国在该领域的投入差距很大，技术发展严重不平衡，部分国家会在一定时期内形成事实上的技术垄断，从而获得大量超额收益，使本已十分严重的全球财富的两极分化进一步加剧。^②

就全球社会经济结构而言，人工智能技术的发展发挥的是“创造性破坏”作用。^③它实际上已经在深刻改变维系国民经济运行和社会生产经营活动的各项基本生产要素的意义。在人工智能技术的影响下，资本与技术在经济活动中的地位获得全面提升，而劳动力要素的价值则受到严重削弱。在弱人工智能时代，传统工业化时代重要的人口红利很可能成为新型经济模式下的“不良资产”。在这种情况下，全球范围内的资本和人才将进一步流向拥有技术优势的发达国家，留给发展中国家走上现代化道路的机遇期实际上会变得极为有限。随着社会经济结构的重大调整，资本的权力也将在政治体系中快速扩张，对建立在传统工业化生产基础上的现代政治体系产生影响，最终在政治权力分配中获得充分的反映。

（二）弱人工智能时代国家权力分配模式的调整

从根本上说，国家权力的分配方式是由社会经济生产方式的特点所决定的，不同时代的生产力水平决定了特定时段最为合理的政治组织模式。威斯特伐利亚体系中的民族国家体制正是目前人类所创造的最适宜工业化大生产经济模式的权力分配方式，因此，当人工智能技术推动社会经济结构逐步变革时，新的社会权力将会随之兴起，国家治理结构与权力分配模式也将做出相应的调整。

首先，资本权力将依托技术垄断地位得到持续扩张。人工智能研究门槛很高，需要诸多基础条件的密切配合。对于深度学习算法的研究来说，需要同时满足三个条件：拥有高性能计算机系统、拥有获取大数据的能力以及聚集足够多的人才。而所有这些条件的获得都有赖于巨额且长期的资本投入，

^① 根据《乌镇指数：全球人工智能发展报告（2016）》的预测，短期内人工智能的主要应用领域将集中于自动驾驶、医疗健康、安防、电商零售、金融、教育和个人助理七个方面。报告全文见：<http://www.199it.com/archives/526338.html>。

^② 参见（英）马丁·福特《机器人时代：技术、工作与经济的未来》，第235—238页。

^③ Mark Montgomery, "Fear of Artificial Intelligence VS. The Ethics and Art of Creative Destruction", <https://www.wired.com/insights/2014/06/fear-artificial-intelligence-vs-ethics-art-creative-destruction>.

这就意味着，人工智能研究中最具实际应用价值的成果多出自大型企业所支持的研究平台。

表-1 全球重要人工智能实验室

公司	名称	主要研究内容
谷歌	AI 实验室	负责与谷歌相关的 AI 产品研发，人工智能系统 TensorFlow 研发
微软	微软研究院	语音识别、自然语言处理、计算机视觉
IBM	IBM 研究院	深蓝与 Watson 项目
Facebook	Facebook 人工智能实验室 (FAIR)	图像识别、语义识别、智能筛选、回答简单问题
	应用机器学习实验室 (AML)	与 Facebook 相关的人工智能与机器学习应用研发
百度	深度学习实验室 (IDL)	深度学习、机器学习、机器人、3D 视觉、图像识别、语音识别等，负责开发百度无人车、百度无人飞行器、百度识图等项目
	硅谷 AI 实验室 (SVAIL)	深度学习、软硬件结合研究
阿里巴巴	阿里巴巴达摩院 (ADA)	消费级人工智能产品研究
腾讯	腾讯 AI 实验室	机器学习、计算机视觉、语音识别、自然语言处理，以及其他与腾讯相关的 AI 产品研发
	优图实验室	图像处理、模式识别、数据挖掘等
	腾讯 AI 西雅图实验室	语音识别、自然语义理解等

资料来源：腾讯研究院、中国信通院互联网法律研究中心、腾讯 AI 实验室、腾讯开放平台《人工智能》，中国人民大学出版社，2017 年，第 74—75 页。笔者根据最新情况有所调整。

目前，几乎所有重要的人工智能领域的突破性成果都是在互联网时代的超级企业推动下出现的，谷歌、微软、IBM、Facebook、百度、腾讯、阿里巴巴这七大互联网超级商业巨头实际上掌控了当前人工智能领域的大部分话语权。如表-1 所示，他们建立的研究机构拥有海量的数据资源，并且拥有大学与其他研究机构无法企及的巨额研究经费，大部分人工智能的顶尖人才都逐渐汇聚到这些超级企业，并且根据企业的目标需求调整研究方向。^①虽然其他独立实验室与小型企业时常也会做出重要的成果，但最终大部分会通过商

^① 人工智能领域的高级科学家由大学向大型企业聚集已经成为一种常态，目前深度学习领域的顶级科学家如杰弗里·辛顿、弗拉基米尔·瓦普尼克、扬·勒·丘恩、莱昂·布托、吴恩达、王海峰、李飞飞等都长期在各大互联网企业任职。而上文提到的七个中美超级互联网企业也因其资本、数据和人才等领域的集聚效应，而被戏称为人工智能领域的七大“黑洞”。

业并购等方式被纳入这七大企业体系之内。人工智能领域研究已经深深地打上了互联网资本的烙印，大型企业对于数据资源以及人工智能技术的控制能力，正在造就他们实际上的垄断状态。掌握数据和算法本身就代表着潜在的权力，这种力量将渗入当代深嵌于网络的社会生活的方方面面，利用算法的黑箱为大众提供希望看到的内容，潜移默化地改变公共产品的提供方式。在弱人工智能时代，资本和技术力量的垄断地位结合在一起，将逐渐稀释传统上由民族国家所掌控的金融、信息等重要权力。

其次，技术权力凭借知识方面的不对称优势在短时间内膨胀。在向弱人工智能时代过渡的时间段内，作为承担人工智能科技开发核心任务的少数精英科学家群体将获得超越常规的影响力和话语权。作为当前人类科技研究的前沿工作，人工智能对从业者的个人素质和学术训练有着极高的要求，掌握关键技术的顶级科学家实际上是一个非常稀缺的群体，他们通过自己的研究工作直接影响社会生产形态的构成。凭借在关键技术方面不对称的知识优势，核心科学家共同体实际上拥有制定新的人工智能时代各项技术标准的权力，整个共同体的研究兴趣在很大程度上决定了技术发展方向与社会经济结构的变迁节奏。^①

在传统的工业化生产时代，科学研究已经被深深地纳入社会生产体系之中，形成了非常稳定的互动关系。由于科技教育的普及，科学家群体的规模较大，流动性也很强，科学家群体很难获得直接的政治影响力。然而，在走向弱人工智能时代的过渡阶段，顶级科学家共同体规模很小，可替代性不强，群体相对比较稳定。更重要的是，人工智能技术所引发的产业革命与社会结构的大规模调整将同时出现，科学家共同体的研究工作本身就成为技术与社会结构变迁的中间环节。他们凭借自己的话语权与影响力将构成一个特殊的技术权力阶层，成为弱人工智能时代社会权力的另一个重要组成部分。该阶层的权力一方面需要在一定程度上依附于传统政治权力和资本权力而存在，另一方面又具有很强的独立性，并对另外两种权力产生一定程度的制约。

^① 关于科学家共同体的概念，可参考（美）托马斯·库恩《科学革命的结构》，金吾伦、胡新和译，北京大学出版社，2003年。

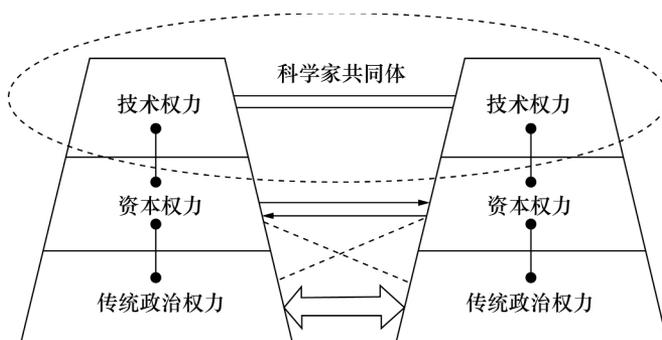


图2 弱人工智能时代行为体内部权力分配模式演变

如图2所示，在弱人工智能时代，传统政治权力仍然是国家内部最重要的权力来源，但资本权力与技术权力的扩张将使行为体内部的权力分配模式发生重大改变。大型企业通过长期投资和技术研发，能够更加经济、有效地在很多领域承担供应公共产品的职能，并从中获取更多潜在的利润。在新的经济形态下，维持旧的公共产品供应格局是一项成本极高的任务，因此，在经济及社会管理等部分领域，民族国家所掌控的传统政治权力会在一定程度上向资本权力让渡公共产品的供应权，并充分利用其在数据和技术方面的优势为政府的行政工作服务，然而，作为这种让渡的代价，传统政治权力会进一步强调自己在军事、安全、金融管理等领域的绝对管辖权，并为资本权力划定扩张边界。同时，政治权力、资本权力也不得不在一定时间内接受技术权力群体的存在，并利用其在相关领域的知识优势推进生产，同时通过各种手段限制其影响力的扩张。

在弱人工智能时代的初始阶段，传统政治权力、资本权力与技术权力将会形成某种金字塔式结构。传统政治权力是塔基，它将继续掌控安全、税收、社会管理等国家权力的核心要素，并继续垄断武力的合法使用权，为其他形式的权力使用背书；资本权力是塔身，大型跨国企业将接受一定领域的权力让渡，为全社会提供数据和经济信息等方面的公共产品，并享有相应的权力；技术权力是塔尖，科学家共同体构成了人工智能技术发展与社会生产的关键纽带，为未来社会提供基本标准和道德影响力，承担推动人工智能技术进一步发展的责任，为资本权力创造利润、为传统政治权力提供产品。三种力量相互支持又相互限制，在不断互动的状态下相互适应，在新的社会发展阶段尝试摸索最为合理的国家治理模式。但三种权力又会因为特性不同，在参与国际互动的过程中表现出不同的特点，因此，在行为体内部权力分配

模式出现变动之后，国际体系层面的互动也必将出现相应的改变。

（三）由内而外的国际体系变革

在威斯特伐利亚体系所确定的国际关系的基本原则中，主权国家是国际体系的核心行为体，直到近几十年国际组织的出现才在一定程度上打破了其垄断地位，然而，在走向弱人工智能时代的过渡阶段，随着资本权力和技术权力在国家行为体内部的扩张，新的异质性行为体将在国际关系活动中发挥更加重要的作用，不同类型的行为体因其不同的特性将带来不同的运行规则及行为逻辑，由内而外地推动国际体系变革的发生。

如图-3 所示，弱人工智能时代的国际体系变革实际上是国内权力分配方式在更广泛空间的自然延展。在新阶段的国际体系中，得到传统政治权力让渡的资本权力和技术权力将获得合法的国际行为体地位，并且将该类型的主体间互动方式引入国际体系中，使整个体系呈现出主体多元、规则多元和关系多元的复杂状态。

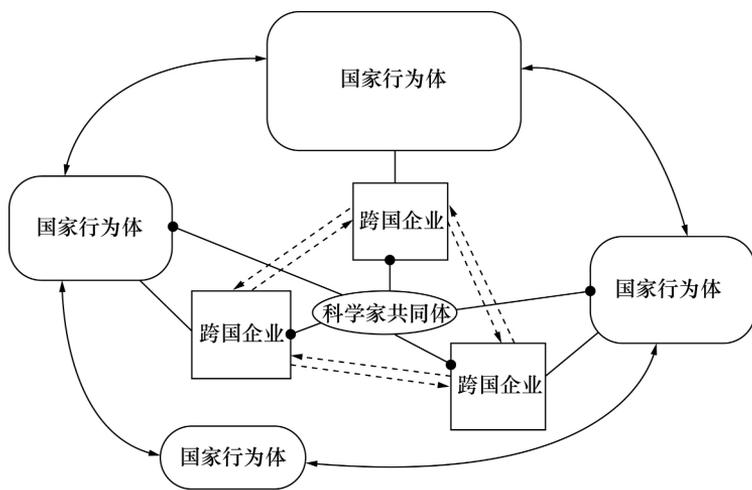


图-3 弱人工智能时代的国际体系

第一，主体多元化。主权国家、跨国企业与跨国科学家共同体都将成为新时代国际体系的重要参与者。科技对于经济运行状态的改变是直接的，而政治体系的变革从根本上说是改变多数人既有观念的尝试。即使在技术进步的情况下，人的观念也常常会因为记忆与惯性迟滞不前。跨入弱人工智能时代，民众与社会也不会在短期内割舍旧时代延续而来的利益纠葛与民族矛盾。因此，传统意义上的国家关系在相当长的时间内仍将延续，主权国家作为国际体系最重要行为体的地位仍难以改变。

然而，由于在相当长的时间内主导着技术开发进程，特别是对大数据资源的独占，使少数大型互联网跨国企业在人工智能领域拥有天然的主导权。数据存储可以通过网络化的形式完成，大数据权力在一定程度上可以脱离主权和地理空间的局限，从而使超大型企业与其所属国家的关系变得十分微妙。虽然主权国家会通过各种方式不断强化自身与跨国企业之间的联系纽带，但对新技术的独占优势赋予非国家行为体在特定公共产品提供方面的特殊地位，而数据与地理的脱钩也使其在相当程度上增强了在特定国际关系领域中的独立性。与人工智能相关的跨国超大型企业，将逐渐发展成以商业利益为核心、具有一定独立性和话语权的国际行为体。他们凭借自身在人工智能领域中的特殊地位为中小国家提供相应领域的公共产品，并以此为基础参与其他议题的政治交易。跨国企业成为新的活跃的国际行为体，将构成弱人工智能时代国际关系的重要特征。

此外，人工智能领域的跨国科学家群体也将在未来国际体系中成为一个独立行为体。科学家群体本身构成了一个拥有共同范式和价值观的学术共同体，学术权威在其中扮演着协调者的角色，如图2虚线圈所示，该共同体既在一定程度上依附于国家和资本，但又有相对超脱的独立运行规则和行为逻辑，在推动技术进步和制定技术标准方面发挥关键作用。该群体人数较少，但具有明显的技术权威与道德优势，对弱人工智能时代全球共识的形成将发挥重要的引导作用。

第二，规则多元化。新时代国际体系中不同行为主体奉行的行为规则存在明显差异。主权国家仍然奉行当前国际社会的基本运行规则，各国仍将围绕权力、利益等传统主题展开竞争与博弈，国家力量仍将是其所依凭的主要手段。而跨国企业行为体的运行规则显然有所不同，由于在企业间关系中没有领土、民族矛盾等顽固政治因素的干扰，经济利益是其竞争与博弈的核心目标，因此，各企业之间虽然也存在重要的竞争关系，但妥协空间相对较大。在人工智能领域崛起的超大型跨国企业往往都在某一细分市场占有绝对优势地位，合作往往可以使其获得较大的额外收益。同时，由于跨国企业本身尚无法获得武力的合法使用权，在与中小主权国家行为体的交往过程中，也主要倡导合作。在该层面的互动中，资本相互合作的动力超过竞争。科学家共同体的特殊性，意味着它必须以科学规则为核心标准才能维持结构的稳定，其运行逻辑具有明显的全球化特征。不同运行规则的同时存在，意味着新时代国际体系的运行注定将是一个复杂的过程，各层次要素需要经过长时

间磨合才能形成稳定的权力边界。

第三，关系多元化。主权国家与跨国企业将围绕权力让渡的边界展开复杂博弈，限制和消解科学家共同体对于技术的垄断，又将成为主权国家与跨国企业共同的政治运作。由于新的行为体和行为规则的加入，弱人工智能时代的国际体系需要通过较长时间的碰撞与磨合才能达到稳定状态。作为新兴的国际行为体，跨国企业需要主权国家在一些特定领域进行权力让渡才能获得自己的活动空间，但权力让渡的边界在哪里，则需要两种行为体通过充分互动来确认。主权国家需要跨国企业，大国需要利用本国所属的跨国企业在技术和商业上的成功带动本国的经济发展和产业的全面进步，中小国家则需要相关跨国企业为其提供必要的服务和技术支持，但主权国家行为体也随时警惕跨国企业的发展，防止其侵蚀主权国家的核心权力、削弱主权国家在政治体系中的主导地位。这一过程将是长期且复杂的，技术进步的速度与社会思潮的变迁都将对双方的博弈产生影响。

然而，在以政治手段消解科学家共同体对于技术的垄断方面，主权国家与跨国公司具有充分共识。由于巨大的知识鸿沟的存在，作为人工智能技术与社会生活的纽带，技术权力阶层的出现是不可避免的，但由人数很少且相对稳定的人群掌握如此关键的社会权力，且无法以正常程序对其核心成员进行更替，在其他任何社会群体看来，都具有相当的风险。由于在人类历史上缺乏处理少数人垄断高阶技术权力的经验，人们也无法保证所有科学家共同体的成员都能够以崇高的道德标准要求自己。因此，如何通过新的机制建设，在充分发挥科学家群体在人工智能领域的积极作用的同时，限制技术权力可能的滥用、努力瓦解科学家共同体，使之分化并依附于各个国家和企业，是主权国家与跨国企业共同的目标，这也将成为弱人工智能时代早期的核心政治运作。

总的来说，人工智能技术的进步对于社会经济结构与权力分配模式的改变，将由内而外推动国际体系走上变革之路。不同性质的新兴权力将通过多元行为体崛起的方式予以充分表达，本文所列出的多元状态是笔者根据人工智能技术的特点进行理论推演得出的结论，但现实世界显然要比抽象的理论推导复杂得多。事实上，对于正处于社会变革进程中的研究者来说，很难预设处于变革临界点的国际关系行为主体会做出怎样的选择，也很难提前设计国际体系合理的重构方案。但可以确定的是，随着持续的技术进步，动荡与不确定的风险会不断累积，国际体系变革将是不可避免的结果，与时代特征相契合的新体系只有在反复磨合之后才能最终形成。

四、弱人工智能时代的思想观念： 保守主义与进步主义的交锋

作为新一轮产业革命的先声，人工智能技术所展现出来的颠覆传统社会生产方式的巨大潜力，以及可能随之而来的普遍性失业浪潮，不仅推动物质与制度层面的改变，也会持续地冲击人们的思想观念。面对剧烈的时代变革与动荡，新的思想理念将会不断涌现，希望为人们所面临的共同问题提供有力的解释和解决方案，最终这些思想与行动将汇集形成具有时代特征的社会思潮，对国家权力分配的变革和国际体系的变迁产生重要影响。

迈向弱人工智能时代，人类社会的观念方面已经展现出两种不同的思想观念体系，这两种思想的差异从根本上说是对人工智能技术的基本认知分歧，也充分反映了人类社会传统思想理念中保守主义与进步主义的分野。^①针对人工智能进步所带来的风险与挑战，他们提出了不同的应对策略，于是，借用政治思想领域的标签，我们可以更加清晰地描述当前人类社会对于人工智能进步的两种基本态度。

（一）保守主义

关于人工智能问题的保守主义路径，其理论出发点是对技术快速进步的恐惧。事实上，在每一次工业革命发生时，人类社会都会严重担忧技术可能带来的风险不可控问题，而这种担心在本轮人工智能的发展问题上表现得特别严重。一方面，这是因为在深度学习算法释放出人工智能技术的发展潜力之后，在很多领域人工智能都仅仅需要很短的学习时间便能够超越人们多年来所积累的知识与技术，人类突然意识到，自己曾经引以为傲的思维能力在纯粹的科学力量面前显得多么微不足道。另一方面，深度学习算法的“黑箱”效应，使人类无法理解神经网络的思维逻辑。对人工智能技术的恐惧，实际上是人类对无法预知的未来世界和对自身力量有限所带来的无力感而形成的双重担忧，这种观念在过去很长时间里通过各种文艺作品被充分表达出来，产生了广泛的社会影响，而保守主义路径就是这种社会思想的集中

^① 关于保守主义与进步主义的思想渊源，可参考刘军宁《保守主义》，东方出版社，2014年；李强《自由主义》，东方出版社，2015年；李颜伟《知识分子与改革：美国进步主义运动新论》，中国社会科学出版社，2010年。

反映。

人工智能领域的保守主义思想强调，要以稳健的态度处理人工智能技术的发展问题，特别是在发展通用人工智能以及超人工智能的问题上表现出极度的谨慎。在保守主义者看来，维持人工智能技术的可控性是技术发展不可逾越的界限。^①

针对弱人工智能时代即将出现的失业问题，保守主义建议利用一场可控的“新卢德运动”^②以延缓失业浪潮，通过政治手段限制人工智能在劳动密集型行业的推进速度，使绝对失业人口始终保持在可控范围内，为新经济形态下的新型就业岗位的出现赢得时间。这种思路的出发点在于尽可能长地维护原有体系的稳定，以牺牲技术进步的速度为代价，促使体系以微调的方式重构，降低整个体系的动荡强度。

但保守主义思想理念自身也存在难以解决的内在悖论。尽管作为一种总体性观念，保守主义能够逻辑自洽，但是，当把这种观念付诸国际体系的竞争环境中时，其理念就很难被接受。在科技快速发展的时代，任何国家放缓对新技术的研发和使用在国际竞争中都是非常危险的行为，人工智能技术的快速发展可以在很短的时间内使国家间的力量差距被不断放大，而信奉保守主义理念的国家将在国际经济和政治竞争中因为技术落后陷入非常不利的境地。因此，这种思想总体上将长期停留在社会民众与知识分子层面。

（二）进步主义

人工智能问题上的进步主义思想，其理论出发点在于承认人工智能技术的进步是不可阻挡的历史潮流，而且也相信科技进步会为人类社会带来积极的影响，主张各国共同加强人工智能技术的研究，利用技术红利所带来的生产效率提升获得更多的社会财富。

进步主义体现了人类对于人工智能技术的向往，这一思想理念高度评价人工智能所引领的本轮工业革命的重要意义。他们解决问题的逻辑是要通过调整制度和社会的基本原则，充分释放人工智能技术发展的红利，在新的社

^① 关于超人工智能的主要特征，可参见 Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 2014。

^② 卢德运动是19世纪初英国工业革命时期由传统纺织业者发起的捣毁机器的群众运动。20世纪90年代后，新一代反对现代技术的哲学思潮逐渐出现，出于对自动化、数字化负面影响的担心，希望限制新技术的使用，由于同早期卢德运动的思想渊源相近，因此其又被称为新卢德运动。在弱人工智能时代，预计新卢德运动将获得更多的支持，有可能成为一股重要的社会思潮。参见 Steven E. Jones, *Against Technology: From the Luddites to Neo-Luddism*, Routledge, 2006。

会原则基础上构建一个更加适应新技术发展特性的人类社会。

在进步主义者看来，人工智能技术所导致的大规模失业是无法避免的历史规律，试图阻止这种状况出现是徒劳的，维持弱人工智能时代社会稳定的方式不是人为干预不可逆转的失业问题，而是改变工业化时代“不劳动者不得食”的分配原则。利用技术进步创造的丰富社会财富，建立“无条件基本收入”保障制度，为全体公民提供能够保障其维持体面生活的固定收入，以鼓励其在长期失业状态下重新获取技能，或从事其他感兴趣的工作，^①最终在新的分配方式基础上重新构建社会文化认知，形成新时代的社会生活模式。

进步主义思想理念的主要矛盾在于，其理论基础建立在人工智能技术能够快速发展并能够持续创造足够丰富的社会财富的基础上，从而满足全球福利社会的需求。然而，人工智能技术的发展并非一帆风顺，从弱人工智能时代到强人工智能时代需要经历多久，至今难有定论。一旦科技进步的速度无法满足社会福利的财富需求，进步主义所倡导的新的社会体系的基础就将出现严重的动摇，甚至会出现难以预料的剧烈社会动荡。

归根结底，弱人工智能时代思想观念的变迁与社会思潮的形成，本质上都是对人与人工智能技术的内在联系这一核心问题的思考。保守主义者对于人工智能技术的恐惧，与进步主义者对于人工智能技术的希冀，分别从悲观与乐观的视角就这种辩证关系进行了阐释。然而，无论如何，我们在思考人类与人工智能技术的关系时始终应该坚信，人工智能技术是人类的造物，是人类知识与理性的伟大结晶。面对人工智能技术突飞猛进的发展进程，人类应该感到欣慰。以深度学习为标志的弱人工智能技术具有明显的工具属性，即便是被视为洪水猛兽的通用人工智能，也同样是人类认识世界和改造世界的工具。人工智能可能给世界带来的威胁远远不及那些人类自己可能创造的恶，我们应该以冷静而客观的态度理解和思考人工智能技术对于世界的影响，在变革且更趋不平等的世界中创造更加稳定、合理、体现人类文明与尊严的体系与制度，这也将是新时代国际关系研究的核心主题。

^① 关于“无条件基本收入”的相关问题，参见 Phillippe Van Parijs, *Basic Income: A Radical Proposal for a Free Society and a Sane Economy*, Harvard University Press, 2017; Karl Widerquist, ed., *Basic Income: An Anthology of Contemporary Research*, Wiley-Blackwell, 2013。

余论：弱人工智能时代中国的战略选择

深度学习算法的广泛应用与人工智能技术的飞速发展改变了传统的工业化路径，引领新一轮产业革命蓬勃兴起。它不仅会推动社会劳动生产率的大幅提升，也将对人类社会的经济生产形态、权力分配模式与国际体系结构产生深远的影响。国际行为体之间的力量差异将进一步扩大，人类将步入一个更加不平等的时代。

为了更好地应对具有时代意义的技术革命与制度变迁，世界主要大国都在积极布局，根据本国的国情制定合理的人工智能产业的发展规划，协调本国在人工智能领域的各种资源，努力占据技术与产业的制高点，并争取在新的国际体系中获得更加有利的战略位置。^①

人工智能所引领的产业革命是中国发展的宝贵机遇，也是中国历史上首次以科技前沿国家的身份迎来的产业革命。作为世界上最重要的互联网大国之一，中国在推进人工智能技术的发展方面具有天然的优势。一方面，中国资本对于人工智能未来的商业前景十分看好，投入了大量的研发资金，聚拢了大量人才。经过 20 多年的发展，以阿里巴巴、腾讯、百度为代表的中国顶尖互联网企业已经成长为国际巨头，他们拥有充足的研究经费，能够在全球范围内搜罗人工智能领域的优秀人才。另一方面，作为世界最大的互联网市场，中国的互联网人口在 2017 年已经超过 7.5 亿人，加之移动互联网的普及，使得相关垂直领域积累了海量的数据，为人工智能技术的发展提供了充足的养料，中国的互联网市场空间足够广大，能够顺利推进研究的商业化进程。更重要的是，中国政府强大的产业规划与政策引导能力能够为人工智能领域的全面发展提供强有力的支持。

^① 2016 年，美国政府先后发布三份关于人工智能的报告，提出了推动人工智能技术发展的全面计划。参见：National Science and Technology Council, Networking and Information Technology Research and Development Subcommittee, “The National Artificial Intelligence Research and Development Strategic Plan”, October 2016, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf; Executive Office of the President, National Science and Technology Council, Committee of Technology, “Preparing for the Future of Artificial Intelligence”; Executive Office of the President, “Artificial Intelligence, Automation and the Economy”。英国政府对于人工智能产业的相关政策可参见：Government Office for Science, “Artificial Intelligence: Opportunities and Implications for the Future of Decision Making”, November 9, 2016, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf。

习近平总书记在党的十九大报告中明确提出，要“推动互联网、大数据、人工智能和实体经济深度融合”，^①将人工智能视为供给侧结构性改革和推动实体经济发展的关键一环。2016—2017年，国务院先后颁布了《“互联网+”人工智能三年行动实施方案》^②和《新一代人工智能发展规划》^③，对中国在人工智能技术和产业方面的战略定位、发展目标、主要任务、保障措施等方面都提出了详细的规划。通过上述规划的实施，推动人工智能科技的发展已经成为我国重要的国家战略，这将有力地推动我国在人工智能技术和产业化方面的进步，也将对全球人工智能技术的发展产生积极影响。在国家规划的推动与协调下，通过对资源的有效整合，我国完全有能力在未来人工智能领域的竞争中充分发挥比较优势，补齐发展短板，在新的国际体系中占据重要位置，实现社会主义现代化强国的建设目标。

最后，我们需要再次强调的是，人工智能革命将给国际体系带来人类历史上前所未有的深层次变革，变革的过程本身就意味着无法预估的风险。如何未雨绸缪，使技术革命的影响不至于反噬人类本身，则是国际关系学所肩负的重要责任。面对人工智能技术所引动的时代浪潮，国际关系学者必须成为一个冷静的观察者，提前思考与设计符合新的时代特征的政治与经济模式，确保在生产方式剧烈变革的时代始终维持国际社会的安全、稳定与秩序。一言以蔽之，在激烈变革的时代探索恒久的人类精神，在渐趋不平等的世界寻求维持和平与稳定的最佳方案，这是国际关系学在人工智能时代所肩负的历史责任，也是每个国际关系研究者不可推卸的崇高使命。

(责任编辑：吴文成)

^① 习近平《决胜全面建成小康社会，夺取新时代中国特色社会主义伟大胜利——在中国共产党第十九次全国代表大会上的报告》，载《中国共产党第十九次全国代表大会文件汇编》，人民出版社，2017年，第24页。

^② 《“互联网+”人工智能三年行动实施方案》的主要内容，参见 http://www.ndrc.gov.cn/zcfb/zcfbtz/201605/t20160523_804293.html。

^③ 《新一代人工智能发展规划》的内容，参见 http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm。