

人工智能时代国家战略行为的 模式变迁*

——走向数据与算法的竞争

封 帅 周亦奇

【内容摘要】 深度学习算法的突破使人工智能技术进入高速发展阶段，技术进步带来的破壁效应推动了国际战略领域的深刻变革。依托持续完善的算法、不断提升的计算能力和海量的数据资源，人工智能技术能够为国际行为体提供更系统的战略评估和精准的战略决策，保证更高效战略动员与执行。随着新技术逐渐渗入战略行为的各个环节，战略领域将先后出现“先发优势”和“数据—算法竞争”两种新的战略行为模式，最终推动人类的战略行为形态从人与人的竞争逐步走向数据与算法的竞争。当然，在战略行为模式重构过程中，将不可避免地遭遇黑箱状态、鸿沟效应等潜在风险。因此我们需要努力降低其中的可能威胁，推动人类社会走向一个更加稳定而繁荣的新时代。

【关键词】 人工智能 深度学习 国家战略与决策 模式变迁 数据 算法

【作者简介】 封帅，上海国际问题研究院助理研究员（上海 邮编：200233）；周亦奇，上海国际问题研究院助理研究员、复旦大学国际关系与公共事务学院博士研究生（上海 邮编：200233）

【中图分类号】 D5 TP18 **【文献标识码】** A

【文章编号】 1006-1568-(2018)04-0034-26

【DOI 编号】 10.13851/j.cnki.gjzw.201804003

* 本文系国家自然科学基金青年项目“人工智能时代的国际关系研究”（18CGJ013）的阶段性成果。

如果某位活跃在 23 世纪的历史学家试图撰写一部人类战略史，那么他一定会惊叹于 21 世纪上半叶人类战略行为模式所出现的巨大改变。

在进入 21 世纪第二个十年之后，随着深度学习算法的广泛应用，人工智能技术迎来了新的高速发展阶段。^① 技术带来的破壁效应对整个国际体系的内涵与行为体之间的互动方式都产生了深远影响，也必将深刻改变战略行为的基本特征。在技术革命的推动下，战略行为模式的变革进程已经悄然启动，人们也将在未来几十年内见证人工智能技术推动新模式诞生与发展的全过程。尽管新模式的最终形态还存在多种可能性，但是人与技术之间的持续互动将最终决定新的行为模式的总体方向。因此，在技术变革方兴未艾之际，国际关系学科的研究应该提前介入，使新兴技术与传统模式的冲击碰撞始终保持在稳定、合理的轨道上。围绕该问题的研究无论对国际关系学科建设，还是对推动国际战略体系的重构，都具有重要意义。

西方国际关系学界对人工智能与国际关系相关问题的实质性研究在 20 世纪 80 年代末、90 年代初便已萌芽，其中以部分美国学者的探索最具代表性。在这一时期，关于人工智能对战略决策过程的影响便是研究关注的重点，并出现了一些有价值的研究成果。^② 但受限于当时人工智能技术的发展层次，有关该问题的研究并未得到国际关系学界的关注。此后，随着人工智能技术发展在 20 世纪 90 年代后期遭遇瓶颈，相关研究也逐渐沉寂。然而，在

^① 关于本轮以深度学习为代表的人工智能高速发展周期的基本技术背景可参考：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*, New York: Basic Books, 2015; Yoshua Bengio, Pascal Lamblin, Dan Popovici, Hugo Larochelle, and Greedy Layer Wise, "Training of Deep Networks," in J. Platt, et al. eds., *Advances in Neural Information Processing Systems (NIPS 2006)*, Cambridge: 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、[加]亚伦·库维尔：《深度学习》，赵申剑等译，人民邮电出版社 2017 年版。

^② 参见 Stephen J. Cimbala, ed., *Artificial Intelligence and National Security*, Lexington, MA: Lexington Books, 1987; Allan M. Din, ed., *Arms and Artificial Intelligence: Weapon and Arms Control Applications of Advanced Computing*, New York: Oxford University Press, 1988; Valerie M. Hudson, ed., *Artificial Intelligence and International Politics*, San Francisco: Westview Press, 1991; Ennals Richard, *Artificial Intelligence and Human Institutions*, London: Springer-Verlag, 1991; George F. Luger, *Artificial Intelligence, Structures and Strategies for Complex Problem Solving*, Harlow, England: Addison-Wesley, 2005; 等等。

新一轮人工智能的关键性技术实现突破后,关于人工智能技术对国际战略的影响的相关研究再次升温。2015年之后,一批有价值的研究成果逐渐问世。^①相对而言,中国国际关系领域针对人工智能相关问题的研究起步较晚,直到2017年才出现与人工智能、机器学习等主题相关的深度研究成果。^②目前在中文的学术刊物和著作中,与人工智能技术和国际战略决策等问题直接相关的研究仍然凤毛麟角,仅有少数几个研究团队对该主题进行了有限的讨论。^③总体来看,国际关系领域围绕人工智能技术与国际战略的研究仍然存在较大的成长空间。

基于此,本文试图对人工智能时代国家的战略行为模式进行深入研究,分析人工智能技术影响国家战略行为的基本路径,展现人工智能时代国家战略行为的模式创新,并深入思考这一过程中面临的机遇、风险与挑战。

一、技术破壁:人工智能介入国家战略行为的基本路径

国家战略行为是指“一个国家为了达到某一目标,在一定计算基础上的行为”。^④它是国家为了应对外部环境的变化,调动并综合运用政治、军事、

^① 参见 Erik Brynjolfsson, and, Andrew Mcfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, New York: Norton, 2014; Martin Ford, *Rise of the Robots: Technology and the Threat of a Jobless Future*, New York: Basic Books, 2015; Jerry Kaplan, *Humans Need not Apply: A Guide to Wealth and Work in the Age of Artificial Intelligence*, New Haven and London: 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, Maaïke Verbruggen, *Mapping the Development of Autonomy in Weapon Systems*, Stockholm International Peace Research Institute, 2017; 等等。

^② 国内相关问题的主要研究成果包括:封帅:《人工智能时代的国际关系:走向变革且不平等的世界》,《外交评论》2018年第1期,第128-156页;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:《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。

^③ 目前,在中国国际关系领域,以人工智能或机器学习为主要研究内容的研究团队数量仍比较有限,主要包括清华大学的庞珣团队、对外经贸大学的董青岭团队、华东政法大学的高奇琦团队、暨南大学的陈定定团队、中国国际问题研究院团队以及由笔者和部分同事组成的上海国际问题研究院团队。总体来说,中国学者对该问题的研究仍处于起步阶段,学术共同体尚未完全形成。

^④ 唐世平、王凯主编:《历史中的战略行为:一个战略思维教程》,北京大学出版社2015年版,第3页。需要说明的是,本文为了更准确说明战略领域的整体变化情况,在对国内外

经济、科技、文化等方面的资源，实现其在安全、发展、荣誉等方面目标的基本原则与方法的总和。^① 经过数百年的实践，国家在战略行为领域构建了较为稳定的运行模式，并已成为国际体系的重要组成部分。

（一）传统国家战略行为的基本模式

战略行为是一项复杂的系统性活动，它包括国家战略的设计过程与执行过程。根据既有的战略理论研究成果，我们可以把国家战略行为分解为相互联系的四个部分。^② 其一，战略评估，即国家对国际环境、国家能力和国家政策选择等因素做出全面、准确评估。战略评估是战略行为的基础。其二，战略决策，即决策者根据战略评估结果确立一整套战略规划或战略政策，用以指导国家政策。战略决策是整个战略行为的核心。其三，战略动员，即战略准备，是国家通过动员国际、国内战略资源（包括物质资源与人力资源），以实践确定的战略。战略动员实际上是国内因素对国家战略行为产生影响的过程。其四，战略执行，即国家利用已经动员的战略资源，按照规划实施既定战略。战略执行也是战略行为的最终表现形式。

如图 1 所示，在现代国际体系形成之后，国家战略行为基本上根据该模式运行。在战略执行过程结束后，根据执行效果重新进行战略评估，重新推动新一轮战略行为，同时，国内体制也会根据战略执行效果加以调整，以便提升动员能力。

各种战略分析模式进行对比之后，选择了唐世平老师提出的“战略行为”概念作为基本分析框架。我们认为这种分析框架能够更好地概括整个战略活动的各个环节，也能够更加全面地展示战略领域的基本状态。

^① 事实上，不同时代的学者对“战略”这一概念均有多重界定，各种概念虽存在外延上的差别，但其核心逻辑基本一致，对本文的研究结论影响较小。对于战略概念的不同界定和应用，可参考李际均：《论战略》，解放军出版社 2002 年版；叶自成：《中国大战略》，中国社会科学出版社 2003 年版；钮先钟：《战略研究》，广西师范大学出版社 2003 年版；杨洁勉：《大合作：变化中的世界和中国国际战略》，天津人民出版社 2005 年版；时殷弘：《战略问题三十篇——中国对外战略思考》，中国人民大学出版社 2008 年版；周丕启：《大战略分析》，上海人民出版社 2009 年版；门洪华：《中国国际战略导论》，清华大学出版社 2009 年版。还包括 Henry Kissinger, *Diplomacy*, New York: Simon and Schuster, 1994; Zbigniew Brzezinski, *The Grand Chessboard: American Primacy and Its Geostrategic Imperatives*, New York: Basic Books, 1997; John Lewis Gaddis, *Strategies of Containment: A Critical Appraisal of American National Security Policy During the Cold War*, New York: Oxford University Press, 2005; 等等。

^② 唐世平、王凯主编：《历史中的战略行为：一个战略思维教程》，第 11-13 页。

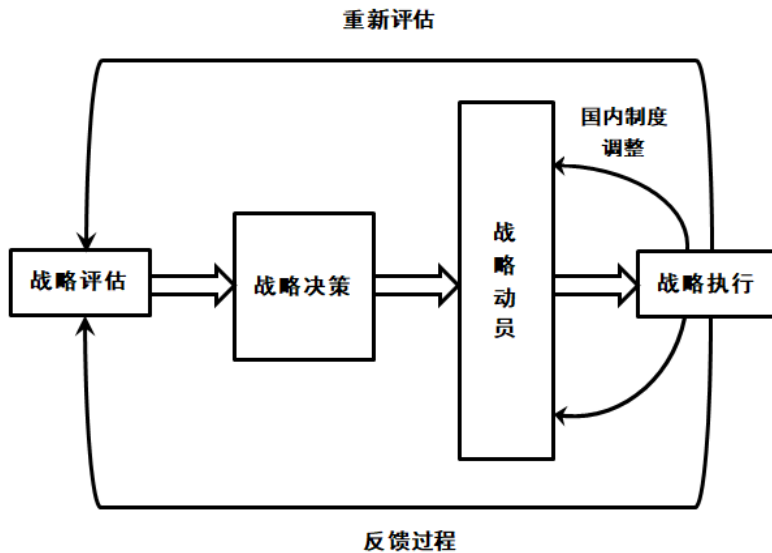


图 1 传统的国家战略行为模式

资料来源：作者自制。

截至 21 世纪初，战略行为每一个环节的工作都是由分工不同的群体承担。例如，信息搜集人员和政策分析人员负责战略评估，思考可能的战略选项，由政治决策层负责做出战略决策，而战略动员与战略执行则更多依靠以科层制为基本特征的管理体系。科层制体系是在第二次工业革命的背景下形成的，是为了适应现代经济制度和大型社会组织而建立的组织和管理系统，充分反映了工业文明的内在要求。^① 随着时代的发展，虽然很多以提高工作效率为目标的新技术已经不断得到应用，但是这一时期的技术进步并没有对战略行为的基本模式造成冲击。历史延续性成为国际战略行为研究的内在核心逻辑，依托历史经验和案例研究的方法依然是战略研究的主流。人们似乎也很难找出基辛格、布热津斯基等当代战略家与中外历史上的战略大师们的实质性差异。历史与习惯构成战略行为传统模式的强大壁垒，这些很难发生重大改变。

^① 参见[德]马克思·韦伯：《经济与社会》（上卷），林荣远译，商务印书馆 1997 年版，第 278-286 页。

然而，科技的进步是打破既有壁垒与改变传统模式的终极武器，当人工智能技术在深度学习算法的推动下进入新一轮高速发展阶段时，技术的“破壁效应”拥有了影响国家战略行为模式的能力，这块人类智慧的“保留领地”开始面临前所未有的变革压力。

（二）人工智能技术介入国家战略行为的有效路径

人工智能技术所引领的本轮科技革命与以往的科技革命有着本质的差别。“过去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到人类改造世界的活动中。”^①如果说以往的科技革命的意义在于使人类的肢体行动能力获得无限延展，那么人工智能技术则将技术发展的目标直接指向了人类的智慧，希望利用机器模拟人类的意识（consciousness）、思维（mind）等大脑的功能，实现改变人类生产、生活实践的目标。

人工智能技术研究从 20 世纪五六十年代开始兴起，其间几经沉浮，始终未能实现大规模的商业化应用。直到 2010 年前后，随着计算机硬件设备的进步和移动互联时代带来的大数据累积，多层神经网络方法再次成为技术进步的关键支点，触发了人工智能领域的“深度学习”（Deep Learning）革命。^②深度学习算法的成熟是人工智能技术的一次飞跃，借助深度学习算法，人工智能技术获得了从海量数据中提取和识别信息，并建立独特的理解问题的框架与解决问题方案的能力。由此，计算机就能够解决涉及现实世界知识的问题，并能做出与人类主观行为类似的决策。^③在深度学习算法的引领下，人工智能的领域性应用范围逐渐扩大，在部分领域的应用越来越成熟，推动人类逐步走近“弱人工智能”时代。国际战略领域则是本轮深度学习算法革

^① 封帅：《人工智能时代的国际关系：走向变革且不平等的世界》，第 129 页。

^② 关于人工智能技术进步的进程，可参考 Ray Kurzweil, *The Singularity is Near: When Humans Transcend Biology*, New York: Viking Press, 2005; Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning*, Cambridge: MIT Press, 2016; 尼克：《人工智能简史》，人民邮电出版社 2017 年版。

^③ [美]伊恩·古德费洛、[加]约书亚·本吉奥、[加]亚伦·库维尔：《深度学习》，第 2 页。

命最可能应用的领域之一。^①

战略领域是一个非常适合人工智能技术发挥作用的领域。一方面，从本质上看，战略领域的活动是一个基于明确对手和稳定目标的博弈过程，这适宜深度学习算法发挥作用。另一方面，由于战略领域在国家政治生活中处于重要地位，丰富的数据资源也有利于深度学习算法作用的发挥。因此，在人工智能技术取得突破性进展后，战略领域就顺理成章地成为技术尝试应用的领域之一。

而对于国家战略行为而言，人工智能技术的参与是对其原有体系的系统性介入。深度学习算法的功能与传统战略行为模式的每一环节都能够实现有效互动，对整个战略行为体系将产生冲击性影响。

一般来说，作为“弱人工智能”时代的标志性特征，深度学习算法所推动的人工智能技术的核心能力主要有四个方面。^②

第一，认知（cognition），即通过机器搜集信息、解释信息的方式来描述世界。主要表现为利用自然语言处理技术、机器视觉技术从图像、声音、视频、文字等方面全方位获取信息，并对海量数据进行分析，针对不同应用场景提供有意义的具体描述。^③

第二，预测（prediction/forecasting），即在获得广泛信息的基础上，通

^① 需要强调的是，由深度学习算法所引领的人工智能技术突破推动了领域性人工智能技术（Artificial Narrow Intelligence）的进步，相关成果只能在特定的垂直领域或应用场景中发挥作用，因此，也被称之为“弱人工智能”技术。目前，该技术距离能够完全模拟人类认知活动，并且实现跨领域、跨场景发挥作用的通用人工智能（General AI）（即“强人工智能”）仍有很大的距离。现阶段我们关于人工智能在国际关系领域的所有讨论都是基于“弱人工智能”时代的技术水平而进行的，这也是科学研究和科学幻想的界限。二者的区别可参见 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.

^② 针对“弱人工智能”时代技术的主要功能，很多研究者都进行了非常有价值的分析与概括，笔者认为麦肯锡公司（McKinsey&Company）的“四功能说”是其中最准确、最全面的概括。相关研究报告参见 McKinsey Global Institute, “Artificial Intelligence: the Next Digital Frontier?” Discussion Paper, June 2017, <https://www.mckinsey.com/~media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx>。以及中文版研究报告：《中国人工智能的未来之路》，2017年3月，http://www.mckinsey.com.cn/wp-content/uploads/2017/03/CDF_McKinsey_AI_CN_final.pdf。

^③ 例如，科大讯飞的语音识别技术、谷歌的智能识图技术都是这一功能的典型成果。

过多层神经网络等方式分析可能出现的不同场景，并且提前预测在各种场景中可能出现的行为和结果。主要表现为在不同场景中预估参与者的行动偏好，在商业和政治活动中则可以针对不同对象的偏好进行有目的的引导。^①

第三，决策（decision-making），即在对搜集到的信息进行有效分析，并完成对特定场景的预测的基础上，根据预设目标确定行动方案。决策是人工智能技术最重要的场景性应用，表现为其有能力在综合分析之后提供最优、次优等不同层次和类型的行动选择，为技术的使用者提供快捷、清晰的多元化行动选项。^②

第四，集成解决方案（integrated solution），即在人工智能技术和其他互补性技术结合之后，可以对极为复杂的活动提供一体化和综合性的解决方案，或建立综合解决平台。这一功能突出表现为借助技术进步，推动既有领域的智能化革新，在不同领域间创造新的联系，形成更有效的行动模式。^③

对于大部分应用场景而言，深度学习算法只需要发挥出某一方面的特定功能就足以改变该领域的商业模式，但战略领域的情况显然更为复杂。在战略行为这一结构复杂、体系完整的场景中，战略博弈的参与方既需要在数据搜集的基础上进行模拟预测，又需要在短时间内、在巨大的压力下做出关键的决策。在此之后，又要在实践层面建立合理机制以整合资源，使策略得到有效贯彻。因此，战略领域对技术应用的需求具有多样化和系统化的特点，这也是以往的技术进步无法改变战略行为模式的深层次原因，直到人工智能技术出现，这种情况才有了改善的可能。

如图 2 所示，在深度学习引领的“弱人工智能”时代到来之际，新的时代前沿技术终于有了全面介入战略行为的有效路径。

^① 例如，淘宝等电商平台针对使用者的消费偏好进行的广告投放行为，以及在脸书平台上出现的针对不同用户的政治类广告投放，都是这一功能的体现。

^② 这一功能的应用范围很广，从优步、滴滴的出行路线规划，电影票的动态定价系统，到 Wealthfront 的资产管理计划推荐，都是人工智能技术的决策功能的体现。而由谷歌旗下的 Deepmind 公司开发的 AlphaGo 系统则是人工智能决策功能最优秀的代表之一，它实际上将围棋这一人类最为深奥和复杂的智力活动分解为多轮决策过程，并最终提供了出色的决策结果。

^③ 例如，Waymo、百度等企业开发自动驾驶技术就是集成解决方案最为典型的案例，其建立基于人工智能的综合解决平台，集成多项领域性技术，满足更为复杂的社会需求。

路径一：人工智能可以通过自身在认知与预测两个方面的强大能力，有效参与战略评估。实际上，传统模式中的战略评估的核心活动是广泛获取与战略目标相关的信息，通过对信息的整理、归纳，建构针对博弈对手的基本分析模型。目前，人工智能技术已经实际参与部分国家的战略评估，这将会成为人工智能在战略领域早期参与的主要突破口。

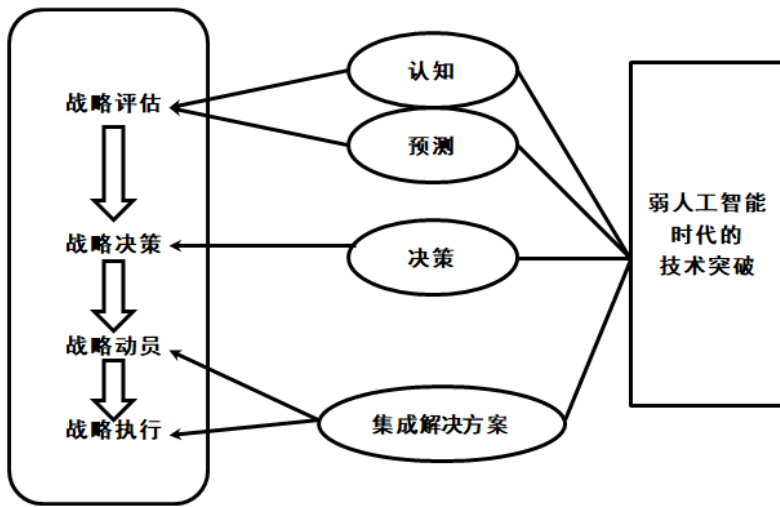


图 2 人工智能技术介入战略行为的基本路径

资料来源：作者自制。

路径二：人工智能在决策方面的强大能力将会为战略决策过程提供更强有力的支撑。作为战略行为的核心环节，如何根据已有的分析进行决策是战略目标能否实现的关键。人工智能决策在该领域与人类决策存在重大差异，突出体现为概率性思维替代确定性思维，众筹式决策替代精英式决策，判断式思维与反复试错性思维的差异。这些差异使人工智能在决策上具备了多线程、强理性和进化型等优势，而这些则是人类决策所缺乏的；新的决策方式缺少道德关怀、灵感直觉等，而这些则是人类决策所独有的。

路径三：人工智能的集成解决方案将会被应用于战略动员与战略执行环节。战略动员与战略执行的过程是调动、分配和使用物质资源的过程。人工

智能与物质力量的结合，如物联网、工业机器人、智能化武器系统等，可以为战略动员提供新的资源，而人工智能技术也将直接缓解战略动员中的集体行动问题，从而使动员更加高效。

综上所述，在人工智能技术通过图 2 所展示的路径，越来越深入地参与国家的战略行为，并发挥越来越大的影响之后，传统的战略行为模式也必然会随之发生改变，并将推动迄今为止人类战略史上最为深刻的模式变迁。

二、数据与算法的挑战：国家战略行为的内生性变革

现实的战略活动是极其复杂的，诸多要素会在战略活动中相互影响，从而产生系统效应（system effects），其复杂程度远远超出人类大脑所能判断和计算的最高阈值。^① 因此，战略博弈往往被视为理性与非理性因素的综合体，是一个涉及理念、意志、政治技巧、勇气、运气等诸多因素的不确定过程。博弈过程往往是在没有完整逻辑链的情况下进行，主要依赖决策者的直觉与判断。但在人工智能技术介入之后，战略行为的内在特性很可能会发生微妙的变化。

在所有具有确定规则和目标的博弈游戏中，深度学习算法都具有相对于人类的明显优势。在充满不确定性的战略环境中，人工智能可以凭借强大的运算性能，更深刻地理解战略领域中各要素的复杂互动状态。战略行为本身可以被抽象为兼具完全信息博弈（complete information game）与不完全信息博弈（Incomplete information game）双重特征的双方或多方的对抗游戏。^② 随着算法的不断完善和数据的不断累积，人工智能系统凭借在信息处理、预测、

^① See Robert Jervis, *System Effects: Complexity in Political and Social Life*, Princeton: Princeton University Press, 1997.

^② 前者类似于围棋、国际象棋等博弈游戏，双方在信息完全开放的情况下进行博弈。随着 AlphaGo 等人工智能系统在围棋上占据压倒性优势后，人类在所有完全信息博弈游戏中已经处于下风。后者则是以德州扑克等游戏为代表，博弈的任何参与者不能完全掌握其他行动者的信息。在德州扑克竞赛中，Libratus 等人工智能系统同样通过“强化学习”的方式在比赛中轻易战胜顶级扑克玩家。现实中的战略博弈更接近非完全信息博弈活动。相关技术性成果可参考 NIPS2017 最佳论文：Noam Brown and Tuomas Sandholm, “Safe and Nested Subgame Solving for Imperfect-Information Games,” <https://arxiv.org/pdf/1705.02955.pdf>。

模拟等方面的强大能力，可以成为人类战略活动的重要助手。当算法优势能够稳定地形成博弈优势时，新技术就会越来越深地嵌入战略行为的各个环节，推动各环节进行自发的适应性调整，最终带来模式的内生性变革。

（一）人工智能技术将提供更加系统的战略评估

全面而快速地搜索信息是计算机相对于人类的常规优势，随着深度学习算法的不断成熟，机器获取信息的方式与渠道得到了极大的扩展，其对信息的分析能力与预测能力也得到了前所未有的提升，从而能够进行更加系统的战略评估。

一方面，深度学习算法将首次实现对战略领域“大数据”的全面获取和深度解读。在现代生活中，海量的日常生活数据是对一个国家综合国力和经济活跃程度的最明显反映，可以清晰地展示一国在国际产业链中的位次、资本的流向、社会思想等情况。在传统战略模式中，由于人类的大脑不足以对如此庞大的数据量进行有效分析，大量数据实际上并不能转化为战略资源，因此传统模式的战略评估只是一种“小数据”范畴的分析方式，从事战略分析活动的人员只能在海量的信息中寻找其中有限的内容进行分析，通过个体推断总体，并以此为基础描绘全局图景。然而，在深度学习技术逐渐成熟之后，战略评估就拥有了大数据时代的技术基础。人工智能在与大数据相关的图像、图片、音频、文字识别方面都有了长足的发展。^① 人工智能技术可以通过“监督学习”和“非监督学习”等方式，使机器在短时间内掌握专业人员所拥有的信息分析和处理能力，能够对海量信息进行分类处理，并以高效率对相关信息进行深度解读，实现真正的自动化分析。例如，人工智能系统可以通过“监督学习”的方式，从数十万条推特（Twitter）信息中自动筛选涉及极端恐怖主义的信息，最大限度提升分析的效率。^②

此外，随着移动互联时代个人生活信息的数字化，人工智能系统还能实

^① 目前以人工智能技术获取信息资料的渠道已经非常丰富，形成了自然语言处理数据集、计算机视觉、图像识别数据集等数十个数据集系统。相关链接汇总参见 <https://www.jianshu.com/p/8c7c1af7d593>。

^② 相关研究案例可参见 Tamar Mitts, “From Isolation to Radicalization: Anti-Muslim Hostility and Support for ISIS in the West,” Working Paper, March 31, 2017, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2795660。

现对作为博弈对手的他国政治精英的个人形象的精准刻画。对他国政治精英和政治领袖的研究向来是战略分析的关键一环。在人工智能技术介入之后，人们可以拓展分析他国政治领导人形象的方式，例如，通过政治领导人在不同场合发表的书面材料及演讲视频、音频信息，人们可以推断其所代表的政治意识，甚至可通过微表情分析，了解其犹豫、不自信的频次，从而更加准确地分析战略对手的关键心理特征。^①

另一方面，深度学习算法能够通过较为精确的沙盘推演和风险预警提供系统而完整的策略方案。以基于代理建模（Agent Based Model, ABM）技术为代表的模拟技术已高度发达，在战略博弈方面的应用也已非常广泛。部分美国研究机构已经成功地利用该技术实现了对文化扩散^②、国际体系结构变迁^③、大选结果预测^④等重要议题的研究，并创建了多种理论模型^⑤。基于现有理论，人工智能系统已经能够在充分获取信息的基础上对有限博弈场景完成沙盘推演。利用人工智能技术，这种模拟可以在短时间内重复进行，在重复模拟次数达到一定标准后，会形成很多人类无法虑及的盲点方案，最大限度地满足全面决策的需要。

人工智能在战略预警方面也取得了突出进展。人工智能的战略预警的内在逻辑是，假定在冲突爆发前，往往会出现很多特定的现象，如抗议示威、小规模冲突、针对特定目标的袭击等。当冲突事件发生的频次超过社会能够承载的阈值后，将爆发具有重大影响的冲突。^⑥深度学习算法可以结合不同

^① Yanchuan Sim, Brice. D. LAcree, Justin H. Gross, and Noah A. Smith, "Measuring Ideological Proportions in Political Speeches," Proceedings of 2013 Conference on Empirical Methods in Natural Language Processing, Seattle Washington, October 18-21, 2013, pp. 91-101.

^② Robert Axelrod, "The Dissemination of Culture: A Model with Local Convergence and Global Polarization," *Journal of Conflict Resolution*, Vol. 41, No. 2, 1997, pp. 203-226.

^③ Lars-Erik Cederman, "Emergent Polarity: Analyzing State-Formation and Power Politics," *International Studies Quarterly*, Vol. 38, No. 5, 1994, pp. 501-533.

^④ 例如，美国 FiveThirtyEight 网站对竞选和体育比赛的预测活动。参见 <http://fivethirtyeight.com/>。

^⑤ 例如，著名学者罗伯特·阿克塞尔罗德（Robert Axelrod）所提出的合作进化战略，就是以 ABM 模型进行的模拟为基础。参见[美]罗伯特·阿克塞尔罗德：《合作的进化》，吴坚忠译，上海人民出版社 2007 年版。

^⑥ 该领域的研究可参考董青岭：《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；Michael D. Ward, et al., "Learning from the Past and Stepping into the Future: Toward a New Generation of Conflict Prediction," *International Studies Review*, Vol. 15, No. 4, 2013, pp. 473-490; Skyler J. Cranmer and Bruce A. Desmarais, "What Can We Learn from

国家的特点，设定动态参数，从该国相关文本、图片、视频、音频等数据来源渠道广泛搜索相应关键词，当相应关键词出现的频度达到特定标准时，就会自动进行风险预警。^①同时，系统还可针对不同危机的特征，提出应对方案供决策者选择。

总之，人工智能技术的不断成熟使战略评估真正进入了大数据时代。以数据累积为基础，战略评估过程可以最大限度地接近以“系统效应”为核心特征的真实战略环境。全覆盖式的策略分析与量化的表达方式，将使后续的战略决策环节从主观决断转变为在不同概率场景中的遴选。战略行为中的主观不确定因素将从第一个环节开始就被持续削弱，更加接近纯粹理性的战略行为新模式的框架也将在此基础建立起来。

（二）人工智能技术将有助于实现更加精准的战略决策

战略决策是战略行为的核心环节，它要求以全面而准确的分析为基础，由决策者最终决定实现战略目标的途径与手段。战略决策本身是一个具有较大风险的行为，决策者要在复杂的利益环境中进行反复比较，并在巨大的压力和紧迫的时间内做出选择。在人工智能技术介入战略决策过程后，情况将会变得有所不同。

第一，人工智能在决策中强调概率性思维（Probabilistic），而人类智能强调确定性思维（Determinist）。战略行为的根本挑战在于战略环境的不确定性，不确定性意味着各种事件是否发生是未知的，各种判断也有被证伪的风险。^②因此，人们一般比较倾向于具有确定性特征的建议与方案。但是对于人工智能而言，不确定性只是一种概率分布函数而已，机器可以提供多种较为有效的方案，并将不同方案的概率用量化的方式呈现出来，从而实现决

Predictive Modeling?” *Political Analysis*, Vol. 25, No. 2, 2017, pp. 145-166; Håvard Hegre, Joakim Karlsen, Håvard Mogleiv Nygård, Håvard Strand, and Henrik Urdal, “Predicting Armed Conflict, 2010-2050,” *International Studies Quarterly*, Vol. 57, No. 2, 2012, pp. 250-270.

^① Neil Johnson, Spencer Carran, Joel Botner, Kyle Fontaine, Nathan Laxague, Philip Nuetzel, Jessica Turnley, and Brian Tivnan, “Pattern in Escalations in Insurgent and Terrorist Activity,” *Science*, Vol. 333, No. 6038, 2011, pp. 81-84; and Håvard Hegre, Håvard Mogleiv Nygård, and Ranveig Flaten Ræder, “Evaluating the Scope and Intensity of the Conflict Trap: A Dynamic Simulation Approach,” *Journal of Peace Research*, Vol. 54, No. 2, 2017, pp. 243-261.

^② 参见 Tang Shiping, “Fear in International Politics: Two Positions,” *International Studies Review*, Vol. 10, No. 3, 2008, pp. 451-471。

策环节的优化。^①

第二，人工智能决策是一种众筹式决策而非精英式决策。在传统的人类战略决策中，决策只是少数精英群体根据自己掌握的信息进行分析。但在人工智能参与决策之后，决策的依据变为海量数据，大批原本无法进入高层视线的基础信息得以挖掘和应用。人工智能决策的数据基础是基层的普通行为体所呈现的广泛信息，代表了一种众筹式的态度。决策依据的变化对整个决策乃至社会心理过程会产生微妙的影响。^②

第三，人工智能决策强调演化性，而人类决策强调固定性。人工智能决策是一个不断变化的过程，可以根据最终结果的变化不断调整其决策的模式，^③ 而人类的判断则相对较为固定，一旦专家对某一个问题的理解和认知模式形成后，就会长期影响其判断。由此可见，与人类相比，人工智能在战略决策中的优势体现为以下几个方面。首先，人工智能技术能够实现更高效的战略决策。在很多战略与军事行动的沙盘推演试验中，人工智能技术都显示出相对于人类博弈对手的竞争优势。这种优势主要体现在两个方面：一是机器具有更快的反应速度，二是机器可以找到超出人类认知范围的重要隐性信息。^④ 反应速度的优势源于人工智能系统的强大计算能力，它能够用穷举的方式在很短的时间内完成人类需要较长时间才能完成的权衡过程。这种速度优势在信息革命的背景下具有更重要的意义，在信息革命时代，决策者与大众在获取信息方面几乎可以同步，留给决策的时间窗口变得非常有限。人工智能技术参与决策之后，能够迅速完成多种决策方案的制定，在时间上能够给人类决策者提供最大的选择空间。寻找认知盲点的优势源于人工智能系

^① 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.

^② 客观而言，众筹式决策与精英式决策孰优孰劣，目前尚无定论。有研究比较了这两种方式在预测冲突发生时的准确度，发现两者都不能令人完全满意，而其中长期在新闻媒体上撰写评论文章的所谓专家给出的判断最为糟糕。Naomi Bosler and Gerald Schneider, "The Oracle or the Crowd? Experts Versus the Stock Market in Forecasting Ceasefire Success in the Levant," *Journal of Peace Research*, Vol. 54, No. 4, 2011, pp. 231-242.

^③ Michael Colaresi and Zuhaib Mahmood, "Do the Robot: Lessons from Machine Learning to Improve Conflict Forecasting," *Journal of Peace Research*, Vol. 54, No. 2, 2017, pp. 193-214.

^④ Kareem Ayoub and Kenneth Payne, "Strategy in the Age of Artificial Intelligence," *The Journal of Strategic Studies*, Vol. 39, No. 5-6, 2016, p. 803.

统的决策方法。由多层次神经网络系统建构的战略决策系统所选择的博弈策略将会是非常特殊的，它并不是对人类决策思路的简单模仿和速度提升，而是在目标导向基础上，系统思考与完成博弈策略和执行流程相关的设计。正如阿尔法狗（AlphaGo）系统不是简单地学习人类的围棋战略，而是可以自主研发新的围棋战略一样，^① 机器可以提供很多人类无法虑及的建议，并可以针对博弈对手的思维盲点迅速调整，从而确保为使用者提供高质量的决策方案。

其次，人工智能技术能够确保战略决策更加理性。由于人工智能决策建立在大数据和概率思考方式上，因此其可以将每个事件发生的可能性进行罗列，从而根据单纯的概率来进行决策，以此保证其决策过程是理性的。在传统战略行为模式中，决策者的主观意识在整个战略行为中扮演着核心角色，决策者不可避免地受到自身认知一致性（Cognitive Concurrence）的影响。在面对大量信息时，人会受到特定文化与社会心理的制约，主动接受和处理符合自身心理预期的证据与结论，形成自己的认知捷径，屏蔽那些与自身偏好不同的信息与建议。^② 然而，在人工智能的决策中，目标、参数与数据是唯一能够影响其决策结果的变量。机器本身不会受到各种情绪化因素的影响，也可以摆脱生理层面因疲劳、压力和集体无意识状态的干扰。^③ 可以说，人工智能决策系统是以一种客观中立、稳妥可靠甚至有些冷酷的思维方式来决策，它是一种接近纯粹理性的思维方式。

^① 关于 AlphaGo 的“进化”，可以比较 Deepmind 团队在两次围棋人机大战之后发表的论文，2016 年，AlphaGo 主要采用 MCTS 蒙特卡罗搜索树进行对弈演算，从人类棋谱中学习。但 2017 年，AlphaGo Zero 通过算法调整，采用上限置信区间算法（UCT），完全抛弃了人类棋谱，进行自对弈强化学习，利用训练好的神经网络，运用博弈树形搜索将胜率在原来基础上提高了近一倍。这些使系统提升的方法很可能被运用到战略博弈中。参见 David Silver et al., “Mastering the Game of Go with Deep Neural Networks and Tree Search,” *Nature*, Vol. 529, No. 7587, 2016, pp. 484-489; David Silver et al., “Mastering the Game of Go without Human Knowledge,” *Nature*, Vol. 550, No. 7676, 2017, pp. 354-359.

^② 此外，人类在战略决策过程中还会受到很多无法预知的情绪因素的影响，例如，一些决策者对面子、荣誉等因素的渴望会直接影响其决策方向；而另一些决策者因性格上的谨慎、怯懦而倾向于夸大风险，使整个决策趋于保守。这些现象都是人类心理的正常状态，但也是非理性战略决策的重要根源。可参考 Richard Ned Lebow, *Between Peace and War: The Nature of International Crisis*, Baltimore: Johns Hopkins University Press, 1984.

^③ 关于身体因素对战略决策的影响，可参见[法]皮埃尔·阿考斯、[瑞]皮埃尔·朗契尼克：《病夫治国》，郭宏安译，江苏人民出版社 2005 年版。

再次，人工智能技术能够在博弈过程中实现系统的自我完善。人工智能在战略决策过程中的另一个重要优势是快速“学习”与“进化”。这是由深度学习算法自身的特点决定的，也是人工智能决策相比人类决策最重要的优势。对所有应用深度学习算法的人工智能系统来说，随着系统在实践中应用的扩大，相关数据量的不断累积，系统能够根据情况的变化不断微调策略，实现自我进化，这也被视为一种强化学习(reinforcement learning)过程。^①在博弈过程中进行学习是基于深度学习算法的战略决策系统的重要优势。博弈的最初阶段是人工智能战略决策系统的优势发挥最不明显的阶段，在这一阶段，所有数据都源于历史经验，博弈对手的策略选择也存在很大的偶然性。但当博弈进程开始之后，双方的互动过程将会快速累积反映现实情况的最新数据，并使人工智能系统根据博弈结果调整算法。同时，根据对方采取的博弈策略，人工智能系统可以不断修正对博弈对手的形象刻画，逐渐理解其思考方式及策略特点，并将这些内容作为完善系统参数的依据。机器系统可以通过算法的调整和计算能力的提升在较短时间内实现系统迭代更新，而人类的学习过程则必须按部就班，且要花费很长时间才能完成。因此，从理论上说，即使人类决策者在博弈之初具备一些优势，也会在人工智能系统的进化过程中很快被超越。^②

（三）人工智能技术有助于保证更高效战略动员与执行

与战略评估和战略决策环节相比，目前人工智能技术对战略执行和动员环节的介入程度还比较有限，积累的经验尚不丰富，但部分初步的尝试结果已经展现出人工智能技术推动该环节变革的巨大潜力。在现有的技术条件下，人工智能技术对战略动员与执行的冲击主要体现在两个方面。

第一，智能化武器的出现将改变战略执行的方式。智能化武器系统的技术开发已经引起了很多国家的重视。智能化武器系统与传统意义上的自动化武器系统的特点不同，自动化武器只是通过建立远程操作系统，实现战场的

^① 参见 David Silver et al., “Mastering the Game of Go without Human Knowledge,” *Nature*, pp. 354-359.

^② 当然，具有优势并不等于机器智能在决策过程中不存在缺陷，机器决策系统也存在一些明显的缺陷，例如，机器决策无法理解模糊性目标，无法虑及政治体系内各种部门利益妥协的问题，纯粹的概率分析容易忽视人类追求小概率冒险行动的可能，等等。

无人化，但智能化武器则是利用深度学习算法，使人工智能系统获得操纵武器的基本技能，通过反复的训练和演习积累实际的战斗数据，最终实现武器的自主作战。^① 自动化武器系统是一个复杂的综合解决平台，集合了自动驾驶、机器视觉、策略构建等多领域的前沿技术，也是对各国人工智能时代技术实力的综合考验。智能化武器协调指挥系统一旦成熟，其效果将远远超过常规武器的作战效果。因为人工智能技术支持多线程任务处理，大批量智能化武器可以基于通用的算法系统进行高效率协调配合，并通过人工智能系统的灵活调整，最大限度获得战场优势。同时，智能化武器系统的出现还能缓解国内战略动员中普遍存在的集体行动问题。^② 由于智能化武器系统能够实现作战人员与战争进程的完全隔离，在有关国家国内受到的政治压力相对较小（但国际伦理压力很可能会增大），这也将改变决策者对使用军事手段的决心，从而使国家对自身战略目标进行调整。

第二，科层制管理体系面临变革压力。新技术的发展同样也会对战略动员环节产生重要影响，最为直接的影响是对科层制管理体系的冲击。在 20 世纪，科层制管理体系成为国家发展的必要因素，也构成了战略动员的主要目标。但在人工智能技术介入之后，科层制管理体系的很多问题进一步凸显。人工智能系统推动的是一种扁平化和网络化管理模式，因为机器智能具有高效的任務处理能力，执行过程可以通过最简单的方式部署与反馈。各项流程都能直接与核心主线对接，并根据战略要求实现网络化协作。超强的核心计算能力使环节更少的扁平化组织体系变为可能，而新的组织形态又将带来更高的任务执行效率。对现有的社会管理系统而言，战略竞争的需求将不断反馈到科层制体系，整个体系都面临着变革压力。尽管这一过程会因制度的惯性与路径依赖变得比较漫长，但只要人工智能技术持续进步，组织方式和管理体系的变革只是一个时间问题。

^① Vincent Boulanin and Maaike Verbruggen, *Mapping the Development of Autonomy in Weapon Systems*, Stockholm International Peace Research Institute, 2017, pp. 27-29.

^② 战略动员本身就是一项集体行动，在任何传统战略模式下，都会出现非常典型的“搭便车”行为和“代理人”行为。而在智能化武器系统逐渐成熟的条件下，国内战略动员在很大程度上将转变为对机器工具资源的调配使用，集体行动问题将会减少。关于这一问题可参见[美]曼瑟尔·奥尔森：《集体行动的逻辑》，陈郁等译，上海人民出版社 2014 年版。

三、“效能”与“黑箱”的悖论：新模式基本形态与潜在风险

人工智能技术的持续进步对传统战略行为模式的影响是根本性的。它依托技术的力量，凭借人类望尘莫及的计算能力以穷举等方式完成精确的战略分析，并实现了对各种方案优势与风险的定量分析，它能够以简单、直接、快速、高效的方式协调执行过程，改变战略行为的特征，使原本更多被视为感觉和艺术的战略行为有可能转变为纯粹的计算与选择过程。面对这即将到来的未来图景，人们由衷慨叹科学的神奇。

随着新技术逐步应用于战略行为中，传统的战略行为模式将不可避免地受到技术变革的冲击。技术与传统模式的互动将是一个较长的过程，在互动的初始阶段与稳定阶段，由于技术介入的深度不同，将会形成两种具有差异的新模式，最终使战略行为走向数据与算法的竞争。当然，模式的变迁也往往带来不可预知的风险。只有当人们更加深入地理解新模式的演进过程、基本形态以及潜在风险之后，才有可能在剧烈变革的环境下维护全球战略格局的总体稳定。

（一）战略行为新模式的形成过程与基本形态

人工智能技术介入战略行为是一个由点到面、由表及里的过程。最初将是少数国家利用自身的技术优势，通过人工智能技术的应用获得不对称博弈优势。在这一阶段，战略行为模式将出现第一次变革，形成一种具有过渡性特征的新模式。随着时间的推移，技术门槛将会逐步降低，与战略行为相关的人工智能技术将不可避免地在全球范围内扩散，在此背景下，战略行为模式也将“进化”，战略行为的内在意义也将深刻变革，最终将迫使人类在“效能”与“掌控”这两种价值之间做出选择，从而形成与传统模式迥然不同的新形态。

如果将新技术的预期能力与对战略行为的有效介入路径结合起来分析，我们可以看到如图3所呈现的模式变迁的情况，即用图表对人工智能技术介入国家战略行为的基本环节进行结构性描述。该图较为清晰地展示了在人工

智能技术持续发展并得以应用的条件下，战略行为模式可能发生的系统性变迁。随着技术介入的程度变化，将出现两轮模式变迁。尽管这两轮模式变迁之间有着明显的联系，但两种模式的核心特质具有本质差异。这也将成为未来国际战略研究的基本背景。

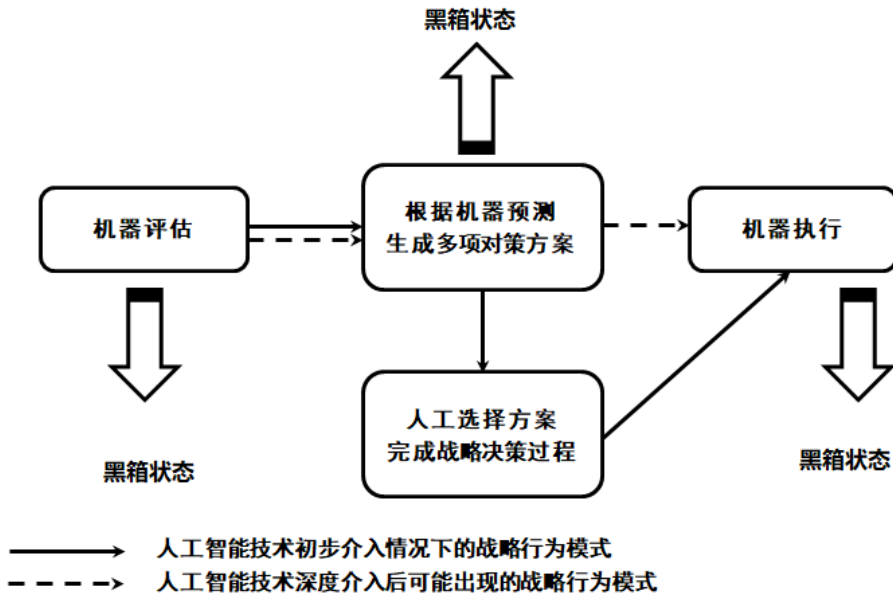


图3 “弱人工智能”时代战略行为的模式变迁

资料来源：作者自制。

新模式一：“先发优势”模式。

如图3实线箭头所示，模式一形成于人工智能技术应用于战略领域的早期，是人工智能技术初步介入战略领域时整个战略行为的基本形态。该模式主要出现在人工智能技术不平衡运用于战略领域的历史阶段。从本质上说，该模式是一种由传统战略行为模式向完全意义上的新模式的过渡，部分率先采用人工智能技术的国家将运用该模式进行战略运筹，与此同时，尚无力应用人工智能技术的国家将继续沿用传统模式。在这一阶段，由于人工智能技术的不平衡应用现象非常明显，可以称之为“先发优势”模式。

“先发优势”模式阶段将是率先应用人工智能技术的少数国家在国际战略领域占据明显优势的阶段。由于重要的技术突破会首先出现在少数科技和数据环境较好的国家，在这一阶段，少数首先掌握相关技术的国家会利用技术进步在博弈过程中获得巨大的比较优势。此时它们的博弈对手仍沿用传统战略行为模式，它们只需将人工智能技术作为辅助手段应用于战略行为中就可以获得明显的战略优势，对技术手段带来的极限效能的需求并不强烈。因此，这一阶段所形成的新模式非常重视维护人与技术在整个体系中的地位差异，维持人类的主导作用和发挥人工智能的辅助作用将成为这一阶段战略行为的两个基本原则。

如图3实线所示，在该模式中，人类将严格掌控战略决策这一关键环节，而在战略评估和战略执行两个环节充分发挥人工智能技术的作用。在这一阶段，人工智能将首先进入传统上由人所承担的战略评估环节，利用计算机系统的信息搜集和数据处理能力进行人类所无法完成的分析与预测，并以精确的方式呈现出利弊明确的策略方案。这也将是早期人工智能技术对战略行为最直接的影响。当然，随着智能化武器及其他物联网技术的发展，人工智能也将逐渐参与战略执行，但应用速度相对会比较缓慢。

需要注意的是，由于这一时期人工智能技术只在少数国家得以应用，先发优势非常明显，也很难出现明显的人工智能技术竞争的局面，对决策速度和执行效率的要求并不处于优先地位，对技术的稳定与人机协调配合的需要会超出对效率的追求，只要充分发挥人工智能在部分环节的优势，就足以满足使用国的基本预期。因此，在人工智能技术初步介入战略行为的情况下，“先发优势”模式将成为部分技术先进国家战略行为的基础逻辑。这种模式将在很长时间内与传统模式同时存在，并且相互竞争，最终推动传统战略行为模式彻底转型。

新模式二：“数据—算法竞争”模式。

随着人工智能技术的发展与成熟，获得技术的门槛将越来越低，与战略相关的人工智能技术在一定范围内逐步扩散将是不可避免的趋势。在人工智能技术广泛使用后，战略行为模式又将出现新一轮变革。如图3虚线箭头所

示，随着人工智能技术对战略行为的介入越来越深入，各国对人工智能技术的使用越来越普遍，部分国家在战略领域的先发优势将受到持续削弱，而国家间的战略博弈也将逐渐向人工智能系统对抗的方向转变。

当人工智能技术被参与战略博弈的双方应用后，战略博弈实质上就转化为主体的数据与算法进行的对抗，这种层面的竞争几乎可以使所有战略欺诈或策略选择方面的制约变得非常有限，战略竞争将会表现为更直接的力量对抗，而且整个博弈过程将会变得更加迅速而激烈。在这种情况下，整体战略行为的相对效能变得异常重要。因此，这一阶段围绕新型战略竞争形态而形成的战略行为新模式可以称为“数据—算法竞争”模式。

如图 3 虚线箭头所示，“数据—算法竞争”模式的核心目的是在各种人工智能技术相互博弈中获得战略优势，通过对流程的简化和机制的调整，实现最大限度发挥人工智能系统能力的目标。因此，整个战略行为将被最大限度地简化为三个环节，在所有环节都希望能够尽可能创造充分发挥人工智能作用的基本条件。人工智能技术除了深度介入战略评估环节外，为了高效，战略决策环节很可能将不再需要经过人类遴选过程，代之以机器自动选取最优方案的方式加以完成。在做出决策后，再由计算机系统完成资源配置，最终推动战略的迅速执行。这种模式的最大的特点是使人工智能在战略行为中获得更大的自主性，通过提升博弈效率获取更大的竞争优势。当然，这也意味着人类在很大程度上放弃了对战略行为整个过程的直接否决权，对于整个社会心理的影响将是极其深远的。

“数据—算法竞争”模式的出现是国际战略领域在竞争环境下出现的进化与选择的自然结果。由于战略博弈对国家生存与发展的意义重大，任何国家都无法承受持续战略竞争失败的后果，一旦人工智能技术在战略领域展现出巨大优势，获取并使用该技术就会成为国际行为体的共同选择。但在技术扩散之后，先发优势消失，博弈的参与者会发现只有充分、彻底发挥人工智能的潜力，才能继续维持既有的竞争优势。效能的提升有助于创造相对优势，而技术失控的潜在风险则并不是效能提升的必然产物，但将是各国面临的共同风险。在这种情况下，在数据与算法的竞争中处于劣势的国家自然会将效

能作为优先选项，而其他博弈对手在激烈竞争条件下只能选择跟进，这实际上推动了战略领域的模式“进化”。^① 只要国际行为体试图参与到这场技术与战略的双重竞赛中，就不得不接受新的模式与规则，这是人工智能技术发展并深度介入战略行为的必然结果，也将构成“弱人工智能”时代战略行为的核心逻辑。

（二）战略行为新模式蕴藏的潜在风险

任何深刻的变革总是与难以预测的风险相伴而生。作为战略发展进程中的重大节点，新模式的建构与系统的转轨也难免面临风险。这些风险既有可能来自深度学习算法本身的特点，又有可能来自过渡阶段体系内因各行为主体间力量对比的失衡而触发的不稳定状态。充分理解战略行为新模式所蕴藏的潜在风险，是深度理解该模式的前提条件，也是维护国际战略体系稳定的理论基础。

第一，“黑箱风险”。本文所讨论的一切关于“弱人工智能”时代战略领域的变化都源于深度学习算法所引发的技术革命。深度学习利用多层次神经网络的基本架构，实现了通过比较简单的方式来完成表达复杂概念的目标，从而解决了影响人工智能技术应用的很多重要问题。^② 如图4所示，在多层神经网络中，每一层的结果都可以使计算机理解一个简单概念，通过不同层级表示桩体的累加，最终形成对非常深刻的问题和理念的认知。但不同层次的神经网络叠加，实际上创造了几乎无法计数的解释路径。所谓“学习”的过程就是通过大数据的持续输入，使数据一次次尝试通过不同路径解决问题，并根据结果给出反馈，使机器自己对每次循环中的变量权重进行微调，最终自主创造更加有效的模式。通过这种方式，我们能够获得更符合需要的问题解决工具，但无法充分解释其解决问题的原理。^③ 对于人类而言，只能

^① 关于国际体系进化的路径问题，唐世平提出的以“变异—选择—遗传”为基本模式的社会演化理论是一种非常具有说服力的研究思路。参见 Tang Shiping, *The Social Evolution of International Politics*, Oxford: Oxford University Press, 2013。

^② 参见：[美]伊恩·古德费洛、[加]约书亚·本吉奥、[加]亚伦·库维尔，《深度学习》，第3页。

^③ 对于不同算法的内生性特点，可参考 Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine will Remake Our World*, New York: Basic Books, 2015。

看到深度学习在各个领域有效拓展，但程序本身始终保持着稳定的黑箱状态。因此，当人工智能技术一步步介入战略行为后，黑箱状态也将必然成为战略行为模式中固定的一部分。（如图 3 粗箭头所示）

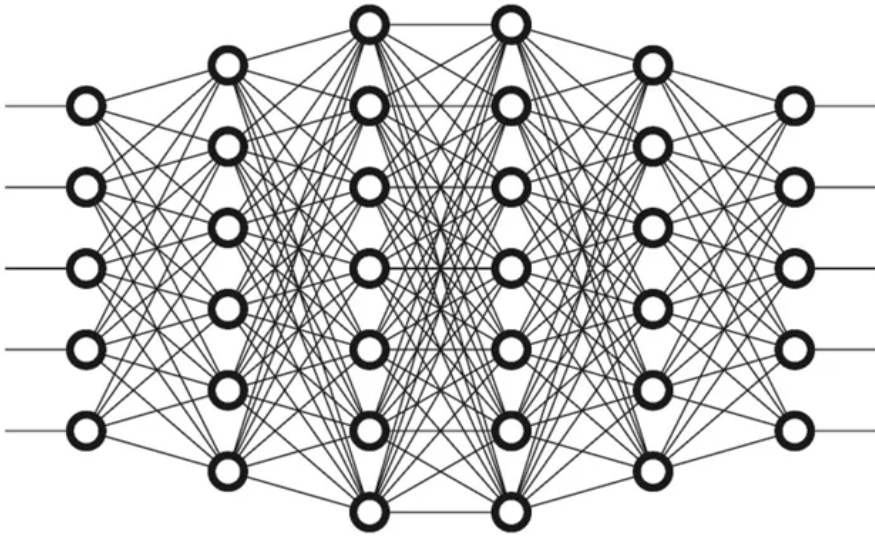


图 4 作为“黑箱”根源的多层神经网络

资料来源：[美]伊恩·古德费洛、[加]约书亚·本吉奥、[加]亚伦·库维尔，《深度学习》，人民邮电出版社 2017 年版，第 4 页。

人在内心深处始终追求世界的确定性与可理解性，而且往往只有在这种状态下社会才能够以较为稳定平和的心态接受科技革命与社会变革的现实。然而，人工智能却是一场注定伴随着黑箱状态不断持续的科技飞跃。因此，无论是决策者还是普通个人，都不可能完全摆脱对人工智能技术快速升级的心理恐惧。

黑箱状态的存在和扩展会带来两方面风险。一方面，由于黑箱状态使人类无法很好地把握人工智能战略推进的节奏与思路，一旦进入“数据—算法竞争”阶段，人类在很大程度上无法对战略实施进程进行及时掌控。对于深度学习算法而言，提前设定系统目标是非常关键的环节。整个算法都会根据

目标进行调整，通过机器学习等方式对系统进行训练，最终形成目标导向的系统性战略安排。然而，在战略领域的实践过程中，除了各种预定的显性目标外，还有可能在战略推进过程中随时出现各种既定目标之外的非意图性目标（unintended impact）。对于战略决策者来说，追求模糊的目标，且不断调整不同目标的战略位次是人类战略智慧的重要组成部分，但这种微妙的模糊性很难被机器所理解，深度学习算法的基本逻辑也很难突破这种系统性困境。而控制这种风险的影响，避免因黑箱状态导致在战术上获益却在总体战略上失败的情况发生，仍然需要依靠人类的智慧。因此，怎样在高度竞争环境下保留人类对整个战略进程的掌控、否定和修改的权利，将成为未来国际关系研究者的一个重要课题。

另一方面，黑箱状态会使整个人类社会对战略行为道德风险产生严重忧虑。以人类为核心的战略行为除了理性的利益评估与计算外，还有内生于人类社会的相应道德、情感标准等行为限制。例如，对特定国家的经济制裁，以不损害普通民众的基本生存权为前提，对正当的军事行动与针对平民的战争犯罪也有严格的区分。但由于人工智能技术的黑箱状态，即使在参数设定上有所考虑，也无法完全保证其依据算法而得出的策略一定能够符合人类的道德情感标准。在巨大的博弈利益与道德准则之间的抉择，将是对决策者内心的严峻考验，也会对人类的道德标准产生强烈冲击。

第二，“鸿沟风险”。人工智能技术深度参与战略行为的另一重大潜在风险存在于国际体系层面。作为新一轮的科技革命的核心技术，人工智能技术的出现会在很短时间内使掌握该项技术的国家与其他国家之间形成鲜明的力量代差。而由于该技术在成熟之前需要长期的基础设施建设、算法研究和数据累积，对资本、人才和市场的要求极高。这也使少数大国与其他中小国家在该领域形成较大的发展差距。在未来相当长一段时间内，人工智能技术及其重要的应用工具都会被少数大国所垄断。

如前文所述，人工智能技术的高速发展将使拥有新技术的国家在战略领域获得巨大优势，而率先掌握新技术的国家也必将是原本就具有较大实力优势的少数大国。缺乏新技术支持的中小国家几乎不可能凭借博弈技巧或经验

弥补原本就存在的实力差距，国家间的战略竞争更容易转变为完全依赖国家实力的对抗。国家间的力量鸿沟将被进一步放大，甚至这种鸿沟最终将难以弥合。^①

人类将在人工智能技术的推动下走向一个激烈变革且似乎更加不平等的世界。^② 这种不平等状况的固化不仅将会影响国际体系的稳定，而且会对国家在新战略模式中的行为方式选择产生负面影响。当弱势一方无法通过传统的、有限的方式对博弈对手构成有效的制衡时，它唯一的选择是打破传统模式中默认的道德界限，以更加极端的方式寻求大国对自身利益的尊重。这是人工智能技术发展带来的“鸿沟风险”的集中体现，各国都将不得不面对一个充满风险的世界。

结论：一场正在进行的“革命”

科技革命正在改变我们所生活的世界，人工智能技术的进步已经使人类触碰到“弱人工智能”时代的门槛，在技术革命的浪潮冲击下，整个人类社会和国际体系都将经历深刻的变革，而战略领域成为迎接这场浪潮的桥头堡。于是，在技术变革的推动下，一场发生在战略领域的“革命”图景正在我们眼前徐徐展开。

随着人工智能介入战略活动的路径逐渐清晰，绵延百年的传统战略行为模式的坚固壁垒在新技术的不断冲击下出现了越来越明显的裂痕。依托计算机的超强计算能力和大数据的持续累积，人工智能技术能够运用相对简单的穷举法，将特定战略议题所涉及的各项影响因素的所有可能变化进行综合分析，并给出明确的方案。在战略分析环节首次实现精确预测，使战略决策转变为在各种精确方案中进行遴选。新技术的介入也将直接冲击原有的科层制管理体系以及现行的资源分配体系，整个战略动员体系将向扁平化和网络化

^① 虽然数据来源不稳定、小概率事件的发生等使人类博弈者在个别场景能够获得优势，但在与人工智能系统的长期博弈过程中，人类取得具有决定性意义的战略优势几乎已经不可能。这也应该成为我们思考未来国际体系变迁的一个背景条件。

^② 关于人工智能技术可能导致国际社会不平等加剧的问题，可参见封帅：《人工智能时代的国际关系：走向变革且不平等的世界》，第 141-142 页。

方向发展。最后，随着深度学习算法向制造业、武器系统等领域的持续推进，加上物联网技术的兴起，战略执行层面的智能化程度也将随之提升。整个战略行为的逻辑都将出现“机器替人”这一难以逆转的趋势，而且越来越倾向于压缩战略行为环节的数量。

新技术的发展能够为使用者在博弈环节带来巨大优势，并极大地提高执行环节的效率，因此该技术将是长期处于竞争性国际体系内的所有行为主体竞争的焦点。尽管存在因黑箱状态引发的道德风险，但任何理性的行为体在获得难以逆转的战略竞争优势这样巨大的“诱惑”面前都不可能选择拒绝。新的技术会因世界各国的竞逐而加速发展，而新的博弈工具和相应的制度安排也将逐步取代传统方案，成为重新塑造国际战略体系的关键力量，最终使整个国际社会在“弱人工智能”时代迎来战略行为模式的变革。

以技术取代艺术，以精确取代模糊，以智能协作取代人际协调，整个战略领域在人工智能技术革命的推动下，正逐步走向算法与数据竞争的新时代。当然，在人类步入一个全新的时代时，与机遇如影随形的总是风险与挑战。在模式变迁中，黑箱状态、力量鸿沟以及其他很多尚无法预知的风险都可能成为新模式的“阿喀琉斯之踵”。因此，对于身处这场变革之中的人类而言，既要理性接受战略行为模式正在改变的现实，并深入把握其变革的基本规律，又要认真反思新模式可能会引发的负面影响以及对体系的潜在威胁，限制并降低新技术发展中蕴藏的不确定性风险，推动人类走向一个更加稳定而繁荣的未来。

[收稿日期：2018-05-14]

[修回日期：2018-05-30]

[责任编辑：石晨霞]