

人工智能军事化与全球战略稳定^{*}

张东冬

【内容摘要】 以人工智能为代表的新兴技术正日益成为影响全球战略稳定的新变量。作为引领新一轮科技革命的战略技术，人工智能在国家安全和军事领域具有颠覆性的潜力。世界主要军事大国已将军用人工智能视为战略竞争新的制高点。人工智能军事化的加快推进改变了传统战争模式以及国际军事竞争态势。在常规武器领域，人工智能技术拥有增强自主武器威慑、参与辅助决策及维护网络安全的能力，可以推动传统战争模式的颠覆性变革，加速战争迈向智能化。在核武器领域，人工智能技术与核导弹预警系统、核指挥和控制系统以及自主核运载平台迭代融合，可以增强核国家核威慑力量，推动核决策走向智能化。从长远角度来看，人工智能技术的军事化应用或将改变现有战略力量的平衡，削弱核国家核威慑的能力，增加危机意外升级的可能性，鼓励军备升级和军备竞赛，冲击和挑战以核威慑力量为基础的战略稳定，进而动摇甚至可能破坏现有全球战略稳定的根基。国际社会应聚焦人工智能军事化过程中相关技术、结构、机制和治理方面的建设，探索降低人工智能军事应用引发战略稳定风险的路径，为建立新型全球战略稳定体系创造有利条件。

【关键词】 人工智能 军事化 核威慑 全球战略稳定

【作者简介】 张东冬，大连海事大学马克思主义学院讲师（大连 邮编：116026）

【中图分类号】 D5

【文献标识码】 A

【文章编号】 1006-1568-(2022)05-0142-20

【DOI 编号】 10.13851/j.cnki.gjzw.202205008

^{*} 本文系中央高校基本科研业务费专项资金资助项目“新时代中国参与全球人工智能治理的理论与实践研究”（3132022319）的阶段性成果。

全球战略稳定事关国际安全大局，历来是国际政治研究的重要问题。传统意义上的全球战略稳定研究大多集中在核武器领域，但随着时代的发展和技术的变革，外空、网络、人工智能等新兴技术正逐渐成为影响全球战略稳定新的变量。^①自21世纪的第二个十年开始，在数据、算法和算力共同进步的推动下，人工智能以前所未有的速度改变其所触及的经济、军事、政治等各个领域，在新兴技术中的革命性地位尤为引人注目。作为技术应用的重要领域，军用人工智能正开启新一轮的军事变革，推动战争形态加速向智能化战争演进，世界主要大国围绕人工智能军事化应用的战略竞争日趋激烈。人工智能技术在军事领域应用的不断强化，深刻影响着传统以核威慑为基础的战略平衡，并给全球战略稳定带来新的冲击和挑战。全面研究人工智能在军事领域的应用与发展，深入考察人工智能军事化与全球战略稳定的关系，探索构建新的全球战略稳定体系，有助于深刻把握全球战略稳定的发展方向，为全球人工智能安全治理及其制度化与规范化提供有益参考。

目前，人工智能与全球战略稳定关系已成为学者关注的重要问题。学术界关于人工智能对于全球战略稳定的影响主要有三种观点。一种观点认为人工智能将冲击和破坏全球战略稳定。^②另一种观点认为人工智能在某种程度上可以增强全球战略稳定，^③还有一种观点认为人工智能并不能对全球战略

① Todd S. Sechser, Neil Narang, and Caitlin Talmadge, "Emerging Technologies and Strategic Stability in Peacetime, Crisis, and War," *Journal of Strategic Studies*, Vol. 42, No. 6, 2019, pp. 727-735.

② 如美国詹姆斯·马丁防扩散研究中心詹姆斯·约翰逊（James S. Johnson）认为，先进常规武器能力的提升将加剧战略核武器和战略非核武器的混合，加快战争的进程和意外冲突的升级，从而破坏战略稳定，增加核对抗的风险；新美国安全中心研究员保罗·沙尔（Paul D. Scharre）指出，诸如机器学习的人工智能应用于自动化，使核预警和指挥控制方面的自主性提高，可能会增加使用核武器的风险，以各种方式影响全球核战略稳定，参见 James S. Johnson, "Artificial Intelligence: A Threat to Strategic Stability," *Strategic Studies Quarterly*, Vol. 14, No. 1, 2020, pp. 16-39; Paul D. Scharre, *Army of None: Autonomous Weapons and the Future of War*, New York: W. W. Norton & Co., 2018, pp. 297-302.

③ 如兰德公司的爱德华·盖斯特（Edward Geist）和安德鲁·洛恩（Andrew J. Lohn）强调，人工智能技术的应用可能加剧核战略不同方面之间的紧张关系，但在有利的情况下也会缓和这些紧张关系，强化战略稳定性；日本国际问题研究所高级研究员户崎洋史（Hirofumi Tosaki）指出，将人工智能引入核武器系统可提高威慑能力的确定性，有助于维持和改善战略稳定，参见 Edward Geist and Andrew J. Lohn, *How Might Artificial Intelligence Affect the Risk of Nuclear War?* Santa Monica: RAND, 2018, p. 20; Hirofumi Tosaki, "Emerging Technologies and Nuclear Deterrence Relationship," *Japan Institute of International Affairs*, April 4, 2021, <https://www.jiia.or.jp/en/column/2021/04/26-emerging-technologies-and-nuclear-deterrence-relatio>

稳定产生根本影响。^①

尽管国内外学者在人工智能与全球战略稳定之间建立了某种联系,但对这一问题的研究仍然存在拓展空间。首先,既有的研究主要建立在现今人工智能技术无法撼动传统核威慑的理论基础之上,探讨的问题也主要集中于人工智能可否影响全球战略稳定,鲜有涉及人工智能军事化如何影响全球战略稳定以及如何建构新的全球战略稳定的研究。其次,与诸多新兴技术一样,人工智能具有发展缓慢的周期性特点,很难在短期内对全球战略稳定产生立竿见影的影响,既有的研究低估了人工智能军事化应用对战略稳定所产生的长期性挑战。再次,即便西方学者在某些方面有所探讨,但大多都是以西方为中心的视角,在很多问题上缺少中国学者的关切和回应。基于对上述问题的考量,结合人工智能驱动战略稳定的革命性和深远性,本文主要探讨人工智能军事化发展及其在常规武器和核武器两个领域中的军事应用,并尝试从长远角度系统评估人工智能军事化应用对以核力量为基础的全球战略稳定可能带来的颠覆性影响,进而尝试探索建构新的战略稳定体系的可能路径。

一、核威慑、新技术扩散与全球战略稳定变化

传统全球战略稳定主要建立在核威慑的基础之上,美苏两国冷战期间关于核力量的博弈构成这一时期全球战略稳定理论和实践的全部内容。冷战结束后,美苏二元核力量体系被多元化核力量格局所替代,特别是随着新兴技术的不断发展和扩散,全球战略稳定理论内涵发生了新变化。

(一) 传统全球战略稳定理论与实践

战略稳定一词起源于冷战时期美国和苏联的核战略竞争,并随着两国核威慑战略的不断演变而逐渐成熟。在冷战的核时代,核武器一度被视为全球战略稳定的基础,大多数关于全球战略稳定理论和政策的形成都与美苏两国

nship.html

^① 如贾子方和王栋认为,人工智能的军事应用有助于形成智能信息化的作战体系,但仅能增强大国的常规战略威慑能力,对核领域战略稳定性不会产生根本性影响,参见贾子方、王栋:《人工智能技术对战争形态的影响及其战略意义》,《国际政治研究》2020年第6期,第55—56页。

战略核力量关系的发展密切相关。1953年，美国国家安全委员会162/2号文件最早提出了战略稳定的模式，即美国与苏联虽然同时具有摧毁对方的核报复能力，但两国却不愿发生战争而形成一种核僵局的状态。^① 20世纪60年代，时任美国国防部长麦克纳马拉（Robert S. McNamara）最先从官方层面提出战略稳定的语义，强调美苏两国在“相互确保摧毁”的前提下，限制和削减进攻性和防御性战略核力量。^② 1990年6月，美苏两国发布的联合声明首次提出战略稳定的一般含义，战略稳定被理解为美国和苏联处于一种战略力量平衡的状态，在这种状态中双方没有发动首次核打击的动机。^③

从理论方面看，战略稳定与核威慑存在必然联系。美国政治学家艾伯特·沃尔斯泰特（Albert Wohlstetter）将战略稳定称之为威慑稳定，认为核反击的平衡有助于威慑稳定性的建构。^④ 正是源于畏惧对手毁灭性的核报复，核国家避免采取大规模的首次突袭，其所形成的稳定态势构成第一次核打击的稳定性。如果一个国家在第一次核打击后仍具有对敌方实施有效的核反击能力，首次核打击能力就会变为第二次核打击能力，这种转化为双方形成新的战略稳定提供可能。核威慑理论认为，战略稳定性来自拥有可生存的第二次核打击能力，在核敌对国家彼此都具有完全摧毁对手的二次核打击能力情况下，各方先发制人的进攻被认为是不必要的，核打击的意愿也会随之下降。也就是说，通过威胁进行全面报复，二次核打击能力可成为核大国关系稳定的前提和基础。

传统意义上的战略稳定有危机稳定性和军备竞赛稳定性两方面含义。危机稳定可解释为敌对国家间没有进行先发制人的核打击的动机。美国核战略学家托马斯·谢林（Thomas C. Schelling）将传统的稳定视为“当危机爆发时，如果双方都没有或未意识到首先出于对另一方使用核武器的恐惧，那么就可

① “Foreign Relations of the United States, 1952-1954,” *National Security Affairs*, Vol. 2, Part 1, NSC162/2, pp. 579-582.

② Robert S. McNamara, *The Essence of Security: Reflections in Office*, New York: Harper and Row, 1968, pp. 61-62.

③ Alexei Arbatov, Vladimir Dvorkin, and Alexander Pikaev, *Strategic Stability After the Cold War*, Moscow: IMEMO, 2010, p. 13.

④ Albert Wohlstetter, “Delicate Balance of Terror,” in Thomas G. Mahnken and Joseph A. Maiolo, eds., *Strategic Studies: A Reader*, New York: Routledge, 2008, pp. 224-242.

以被定义为危机状态下的稳定”^①。在危机稳定的状态中，无论是由于相互威慑、相互信任，还是其他方面原因，对抗各方均无先发制人动用核武器或促使危机升级的动机。^② 军备控制稳定性是在数量和质量上限制发展战略核力量的动机。根据军备控制理论，军备控制的主要任务是调整军备格局，由此减少潜在对手间爆发核战争和出现军备竞赛的可能。^③ 通过塑造攻守平衡，军备控制可以增强危机稳定性，缓解安全困境，进而加强威慑力并最终防止战争。^④

从实践方面看，战略稳定主要基于美苏双边核威慑的战略博弈。冷战初期，美国核力量对苏联具有垄断优势，美苏两国处于一种战略非平衡状态。其间，艾森豪威尔政府曾提出大规模报复战略，企图利用美国核垄断优势抵消苏联的地区常规力量优势。20 世纪 50 年代中期以后，苏联研制出洲际导弹并成功发射人造卫星，获得直接摧毁美国本土的远程核打击能力，美苏核战略态势发生根本性变化。美苏两国同时具备第一次核打击能力，意味着双方都能够将对手的核力量消灭殆尽或摧毁到己方防御系统能够承受的地步。苏联核力量的快速增长使美国对苏联可能采取先发制人核打击的不安全感急剧上升。在所谓“导弹差距”态势下，美国通过加快战略核力量的建设，试图强化战略核武器优势，导致美苏之间的核军备竞赛愈演愈烈。

至冷战中后期，随着苏联战略核武库不断扩张，美国单方面保持的核优势不复存在，美国和苏联同时拥有可靠的二次核打击能力，并形成相互威慑的恐怖平衡，双方都有能力进行有效反击来稳定两极间的对抗。从核威慑角度看，二次打击的核报复能力愈强，美苏之间发生核战争的概率也就愈小，反而有利于达成更为稳定的核均衡关系。因此，美苏两个超级大国在建立各自军事力量之间的稳定平衡、威慑关系和避免核战争方面拥有既得利益，减少核军备竞赛的动机或发动先发制人的核打击，成为两国维持全球战略稳定

① Thomas C. Schelling, *The Strategy of Conflict*, Cambridge: Harvard University Press, 1960, p. 207.

② Thomas C. Schelling, “Confidence in Crisis,” *International Security*, Vol. 8, No. 4, 1984, p. 56.

③ 李彬：《军备控制理论与分析》，北京大学出版社 2006 年版，第 84 页。

④ John D. Maurer, “The Purposes of Arms Control,” *Texas National Security Review*, Vol. 2, No. 1, 2018, p. 13.

的基本守则。特别是古巴导弹危机发生后，美国和苏联就诸多武器技术进行了军控谈判，并采取一系列合作措施，使全球战略稳定性得到进一步加强。

（二）全球战略稳定内涵发生新变化

后冷战时期，世界进入到“第二核时代”^①。由于以冷战为特征的东西方核二元对立不复存在，拥有世界上最大核武库的美国和俄罗斯通过军备控制来寻求稳定的政治意愿逐渐减弱，冷战期间建立的军备控制框架日趋瓦解。随着国际战略环境变化以及世界秩序向多极化演变，以美苏为中心的核战略体系被全球核多元化格局打破，除了五个主要核国家外，以色列、印度、巴基斯坦、伊朗及朝鲜等受区域对抗环境影响，先后拥有研发和生产核武器的能力，事实上或潜在地成为有核国家。同时，冷战后的核战略环境也变得愈加变幻莫测，个别国家在追求发展核武器过程中无视国际法或国际规范，极端恐怖组织等非国家行为体试图掌握小规模核武器技术，这将突破传统核威慑的战略平衡，使国际核战略态势发展呈现出多样化和复杂化的特征。^②

在核武器横向扩散的背景下，全球战略稳定也逐渐超出传统核理论范式。在传统危机稳定和军备竞赛稳定狭义定义的基础上，全球战略稳定同时被描述为有核国家之间没有武装冲突，或者更为广泛地解释为各国享有和平与和谐关系的区域或全球安全环境。^③就俄美相互核威慑的持续性而言，影响全球战略稳定的主要因素包括战略进攻性武器的生存能力、反弹道导弹系统、第三国的核武器、战术核武器、常规精确制导武器、太空武器、反潜战等各个方面。^④从更大的范围来看，全球战略稳定已经突破传统意义上大国间的核战略关系，可理解为全球主要国家的政治军事关系以及全球和平稳定的安全环境。^⑤尽管大国间的核战略关系对于全球战略稳定依然重要，但核

① Colin S. Gray, *The Second Nuclear Age*, Boulder and London: Lynne Rienner Publishers, 1999, pp. 1-2.

② Linton Brooks, Francis J. Gavin, and Alexei Arbatov, *Meeting the Challenges of the New Nuclear Age: U.S. and Russian Nuclear Concepts, Past and Present*, Cambridge: American Academy of Arts and Sciences, 2018, p. 13.

③ Elbridge A. Colby and Michael S. Gerson, *Strategic Stability: Contending Interpretations*, Carlisle Barracks: U.S. Army War College Press, 2013, pp. 117-118.

④ Alexei Arbatov, Vladimir Dvorkin, Alexander Pikaev, and Sergey Oznobishchev, *Strategic Stability After the Cold War*, Moscow: IMEMO, 2010, p. 18.

⑤ 张东冬：《不对称互动、系统性错误知觉与中美战略稳定关系》，《世界经济与政治论坛》2020年第3期，第1—2页。

武器已不再是影响战略稳定的唯一且全部的要素。

实际上，核时代以来全球战略稳定理论内涵的变化主要归因于技术的进步。人工智能、遥感、网络、高超音速飞行器、增量制造、隐身、精确制导等领域的重大进展，让世界再次处于一个新技术时代的边缘。^① 技术的进步及地缘政治环境的变化扩大了威慑的外延，更多地强调地区稳定、常规威慑和多域战争。除了核力量外，威慑的稳定性也越来越依赖于新技术所带来的不断增强的常规能力。一方面，以机器人、无人自主系统、新型高超音速导弹等为代表的新兴技术的发展，加快了非核常规武器的研发和应用步伐，深刻影响了大国未来战略互动方式；另一方面，新兴技术的发展也引发了核技术的变革，特别是核力量的生存能力受到越来越多的威胁，深刻改变了传统核威慑理论的基础。美国学者克里斯托弗·奇巴（Christopher F. Chyba）认为，“新的技术具有直接和间接的军事意义，在某些情况下可能威胁到战略稳定，成为影响战略稳定的重要因素。”^②

二、人工智能技术及其军事化应用

新兴技术的快速进步具有重大军事意义，追求新技术的国家有强烈动机利用这些技术来增强对国家安全极为重要的军事能力建设。在所有新兴技术中，人工智能的革命性地位和影响最为突出，人工智能技术的军事化应用已成为全球主要军事大国竞相进行战略部署的关键领域。

（一）人工智能技术军事化的特点与趋势

人工智能军事化的核心是可执行具体智能任务的技术在军事领域的应用，其目的在于提高传统武器性能及优化传统指挥决策。作为特定领域的一项应用，人工智能军事化具有四个特点。一是颠覆性。人工智能技术可推动以往传统战争逐步向智能化、无人化、网联化迈进，通过赋能武器装备升级、

^① Todd S. Sechser, Neil Narang and Caitlin Talmadge, “Emerging Technologies and Strategic Stability in Peacetime, Crisis, and War,” *Journal of Strategic Studies*, Vol. 42, No. 6, 2019, pp. 727-728.

^② Christopher F. Chyba, “New Technologies and Strategic Stability,” *Journal of the American Academy of Arts & Sciences*, Vol. 149, No. 2, 2020, p. 150.

催生新的战争形态和作战样式、改变战争制胜机理，对信息化战争产生颠覆性影响。二是不平衡性。目前人工智能军事化的应用仅局限于主要军事强国，美国和俄罗斯等大国在人工智能军事化研发和应用方面拥有领先优势，而中小国家因技术能力的差距而相对参与有限。三是风险性。人工智能技术发展缓慢且不成熟会致使军用人工智能面临一系列不可靠和不安全的风险问题，国际机制和规范的缺失也将导致人工智能军事化应用面临法律、伦理和道德方面的风险。四是非透明性。人工智能技术涉及国家战略发展的重大利益，在军事领域的测试、评估和验证具有一定的独立性和保密性，缺乏透明度和实际应用经验。

此外，人工智能军事化与人工智能技术本身的发展密切相关，有渐进发展的周期性特点。在本轮人工智能浪潮兴起的背景下，人工智能军事化发展势头已呈现不可阻挡之势，全球各国也在顺应时代潮流推进人工智能军事化，但其对全球的影响并不会立竿见影。按照事物发展的客观规律，人工智能军事化与人工智能技术本身一样，也将经历从低级到高级、由弱到强、自初创到成熟的不同发展阶段。目前，人工智能技术军事化正处于弱人工智能阶段，仅局限于解决特定领域的具体问题，甚至很多技术难题还未能及时解决，诸多现有技术缺陷亟待完善。但从长远来看，随着技术不断走向成熟，人工智能军事化的步伐将会明显加快，未来的人工智能军事化也将向强人工智能的方向发展，世界主要军事强国有望建立起更成熟、更强大的人工智能军事化应用体系，对全球军事化变革产生更加复杂和深远的影响。

（二）人工智能技术渗透于常规武器

人工智能技术军事化在常规武器中的应用包括自主武器系统、辅助指挥决策以及网络安全行动，这些军事化应用推动传统战争形态的颠覆性变革，加速战争向智能化迈进，正日益发展为与核武器一样的战略威慑因素。

第一，增强自主武器威慑。作为可独立选择和攻击目标的武器，自主武器能够在人类有限参与或不参与的非结构化环境中选择并攻击目标。^① 早期军用自主智能武器通常在人类的直接控制下操作，但目前无人系统已独立于

^① International Committee of the Red Cross (ICRC), *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects*, Geneva: ICRC, 2014, p. 5.

人的干预和监督而走向更大自主，可执行一系列核心空中拦截、两栖地面攻击、远程打击和海上行动。在过去十年中，自主武器系统的研发、部署和应用已经受到世界军事强国的高度重视。如美国军方创建了大量的无人系统，MQ-9“死神”无人机和福斯特—米勒“魔爪”后勤作业机器人曾被部署到阿富汗和伊拉克的战场。俄罗斯也不断致力于在军事行动中使用无人系统，“Uran-6”地面作战机器人曾投入到叙利亚战场，“Platform-M”无人侦察车已在俄罗斯太平洋舰队服役。此外，以色列、英国、法国、韩国、印度等国家也纷纷加入智能导弹、无人机、机器人哨兵的研发和部署行列，试图尽快实现武器系统向自主化和智能化的转变。

第二，参与辅助指挥决策。人工智能系统具有较强的算法处理能力，能够提升军事决策质量速度及战场态势感知。首先，人工智能指挥和控制系统能够避免人类在战争迷雾决策中的许多固有缺陷，减少“人为因素负担”，提高可操作的军事情报和军事决策能力。^①其次，人工智能的机器学习可以在早期侦察到可能存在的安全威胁，同时在环境恶劣的条件下感知并获取相关数据，为增强指挥官态势感知能力和判断准确度提供信息。再次，配备人工智能的早期预警系统可对突然的袭击提出准确警告，提高先进核预警系统的信息处理速度和质量，给决策者留出更多时间进行反应。如美国各军事部门就已经将人工智能纳入无人系统中，用来感知环境、识别目标、传输数据及规划导航，并运用“深度学习”技术训练出具有逻辑分析能力的机器，为决策者提供有效咨询。^②

第三，维护网络空间安全。作为推进网络空间行动的关键技术，人工智能可增强现有的网络防御能力，有效应对敌方的网络攻击和网络干扰。在网络防御战中，人工智能技术可以识别网络行为模式的变化，检测网络异常和软件代码漏洞，并用诸如“深度学习”之类的机器学习技术来识别正常的网络活动，从而对网络入侵形成更强大的防御。在网络攻击战中，掌握先进人

① James S. Johnson, “Artificial Intelligence and Future Warfare: Implications for International Security,” *Defense & Security Analysis*, Vol. 35, No. 2, 2019, p. 4.

② 周琪、付随鑫：《美国人工智能的发展及政府发展战略》，《世界经济与政治》2020年第6期，第40页。

工智能技术的国家利用网络进攻武器，可以迅速搜寻和发现对手隐藏地点和系统漏洞，或者利用机器学习生成的“深度伪造”技术致使敌方生成错误感知，以期获得对相关系统和设施进行先发制人打击的优势。此外，在电子战中应用人工智能技术也可用来干扰对手的目标传感器和通信，通过采取迷惑、欺骗和干扰等手段，导致敌方信息获取、处理及传输的失败，削弱甚至摧毁其攻防力量。如俄罗斯军方就曾在包括叙利亚和乌克兰东部在内的战区部署电子干扰机用于对抗带有全球定位系统（GPS）导航的无人机。^①

（三）人工智能与核武器的技术融合

除应用于常规武器，人工智能技术也与包括核导弹预警系统、核指挥和控制系统及自主核运载平台等诸多方面核力量相融合，增强拥核国家核威慑力量，进一步推动核决策智能化。

第一，人工智能技术增强核威慑力量的有效性。传统“三位一体”（Nuclear Triad）核力量由陆基洲际弹道导弹、潜射弹道导弹和战略轰炸机核力量组成。随着人工智能技术在机器学习及深度算法方面取得重大突破，并加速向军事领域转移，传统的核力量也不断呈现智能化态势。在技术变革推动下，美国和俄罗斯正在考虑使用无人系统进行核运载的可能性。如美国新型战略轰炸机“突袭者”（B-21 Raider），可在无人操作的情况下飞行，并具有执行核任务的能力；^② 俄罗斯核动力无人潜水器“海神”（Poseidon）可携带 200 万吨当量的核弹，从近岸、岛礁和港口的预置装置发射，或在近海千米以下引爆并掀起巨大的海啸，打击和摧毁对方的航母、海军基地和沿海城市。^③ 另外，基于人工智能技术的新型战略核武器带来了核武器投放和核攻击、核防御的新方式。如俄罗斯高超声速滑翔飞行器“先锋”（Avangard）既能投送核武器，也能运送常规武器，可极大地减少美国的指挥和控制预警或响应时

^① James S. Johnson, “Artificial Intelligence and Future Warfare: Implications for International Security,” *Defense & Security Analysis*, Vol. 35, No. 2, 2019, p. 6.

^② John A. Tirpak, “The Raider Takes Shape,” *Air Force Magazine*, Vol. 102, No. 12, 2019, pp. 38-42.

^③ Mark Episkopos, “Russian Navy Will Soon Deploy 32 ‘Poseidon’ Nuclear Drones Across 4 Submarines,” *National Interest*, January 15, 2019, <https://nationalinterest.org/blog/buzz/russian-navy-will-soon-deploy-32-poseidon-nuclear-drones-across-4-submarines-41617>.

间，威胁美国核威慑力量的可信度。^①

第二，人工智能技术在核决策中扮演重要角色。全球人工智能正进入新的快速增长期，技术的应用在核决策中也在扮演愈加重要的角色。由于越来越多诸如卫星监控器、地面雷达和监视飞机等传感器将情报输入作战管理系统，指挥官可能被有关敌方行动的信息所淹没，无法迅速作出明智的决策。人工智能可以通过闪电般的速度筛选传入的数据并突出显示最重要的结果以及区分核攻击的错误警告和真实警告，能够帮助克服上述这些挑战。^②与此同时，具有人工智能功能的决策支持系统，可以帮助指挥官和最高决策者对确认的敌军攻击迹象作出核反应和非核反应，缩小对敌方行动的决策时间，并为指挥官提供最佳选择的算法。核国家还可以依赖复杂的指挥、控制、通信和情报系统的强大能力，定位和摧毁对手的核力量，并实施导弹防御行动。^③鉴于人工智能技术的不断发展，美国、俄罗斯等核大国已经开启推进核指挥控制系统现代化和自动化升级的进程。

三、人工智能军事化对全球战略稳定的影响

与人工智能军事化应用相一致，人工智能军事化对战略稳定的冲击和影响也主要表现在常规武器和核武器两个不同应用层面。从长远角度来看，人工智能军事化可能衍生新危险和新风险，冲击和挑战传统以核威慑为基础的战略稳定，进而动摇和破坏现有全球战略稳定的根基。

（一）改变现有战略力量的平衡

作为颠覆性技术，基于人工智能的自主武器可能会像核武器一样从根本上快速打破现有军事力量平衡，降低发动战争的成本，增加战略不稳定的风险。在人工智能军事化时代，自主武器系统使一个特定任务所需的作战人员

① Steve Trimble, "A Hypersonic Sputnik?" *Aviation Week*, January 14-27, 2019, pp. 20-21.

② Kelley M. Saylor, "Artificial Intelligence and National Security," CRS Report, R45178, November 21, 2019, pp. 12-29.

③ James M. Acton, "Escalation through Entanglement: How the Vulnerability of Command-and-Control Systems Raises the Risks of an Inadvertent Nuclear War," *International Security*, Vol. 43, No. 1, 2018, p. 59.

变得更少，让战斗延伸到以前无法进入的区域，与之相应的军队进攻能力和战斗力也将提升数倍。^① 此外，人工智能技术所提供的战场实时信息和分析，可极大提高国家在陆地、海洋、空中、太空和网络上运行自主系统的执行力和杀伤力，并改变冲突升级的模式，打破全球战略稳定既有的动态平衡。

与核武器较高的研发门槛有所不同，自主武器等人工智能技术对于全球所有军事力量而言，具有易获取且价格低廉等优势。^② 人工智能技术的自由流动可以扩大国家战略竞争的范围，对现有的战略平衡构成破坏性威胁。比如，较小规模的核武器国家可能无法与规模较大的核武器国家相竞争，但它们有可能发展有效的常规武器的能力，使用无人机或智能导弹将军事资产投入天空。一旦掌握人工智能技术的较小规模国家拥有相对较强的竞争力，这些国家在面临大国威胁时对抗的意愿可能就会增强，双方发生冲突的概率也会增加，进而影响全球战略稳定的大局。

自主武器系统还有可能落入包括极端恐怖组织等非国家行为体之手，影响地区和国际力量平衡及战略稳定。自主武器轻小、廉价且缺乏有效监管，为其流入灰色地带提供了便利。目前，“伊斯兰国”等极端组织已经在战场上部署和使用了武装无人机。由于极端恐怖组织可以获得它们自己无法开发的相对复杂系统，那么除了以精确的方式进行有针对性的暗杀之外，还会不加区别地增加自主杀伤的能力。^③ 从某种意义上说，廉价武器化的人工智能增强自主系统的扩散，可能会降低未来无人机攻击的门槛，对地区和国际安全稳定造成严重威胁。

（二）削弱核国家核威慑的能力

人工智能在军事领域的快速应用可能会打破常规战略武器与核武器的边界，通过“以常替核”或“以常制核”削弱核国家既有的核威慑力量。一些国家可借助人工智能新技术，使用常规武器以最快速度定位、追踪并精准

^① Gary E. Marchant et al., “International Governance of Autonomous Military Robots,” *Columbia Science and Technology Law Review*, Vol. 7, No. 12, 2011, p. 275.

^② Future of Life Institute, “Autonomous Weapons: An Open Letter from AI & Robotics Researchers,” July 28, 2015, <https://futureoflife.org/open-letter-autonomous-weapons/>.

^③ Jürgen Altmann and Frank Sauer, “Autonomous Weapon Systems and Strategic Stability,” *Survival*, Vol. 59, No. 5, 2017, p. 127.

摧毁敌对国家传统意义上安全可靠战略核设施。^① 人工智能还可以通过扩展的数据集进行预测，并定位、跟踪和瞄准战略导弹，如地下发射井、隐形飞机上的移动洲际弹道导弹发射器和战略核潜艇中的战略导弹。^② 人工智能对核威慑最不稳定的影响，将是用一系列机器学习改变以往核威慑力量安全可靠的生存方式，削弱有核国家对二次打击能力生存的信心。

不论是有核国家还是无核国家，拥有先进人工智能技术，意味着通过使用常规武器解除一个国家核威慑力量能力的极大提升，增强第一次打击能力有效性，并降低对手的第二次打击能力。正因如此，人工智能技术定位和打击对手战略资产的能力可能会破坏战略核威慑的基本原则，使其战略核力量的生存能力暴露于危险之中，并将核威慑动机从接受彼此的弱点转变为利用对方弱点，形成“单方面确保摧毁”局面。即便没有先发制人攻击的意图，拥有先发制人能力也会成为一些国家设法追求的目标，虽然这样仅仅增加彼此之间讨价还价的筹码，但也将对战略稳定产生消极影响。^③

更为严重的是，拥有智能化常规武器国家具有的执行战略任务的强大能力，其常规行动不论是直接攻击对手的核资产，还是攻击诸如警报或指挥控制系统的两用设施，都可能造成对方压力的不断升级，诱使其采取更加激进的战略行动来增强自身战略生存能力。^④ 在人工智能时代，由于人工智能技术高速度、精准性和无偏见的特征，核国家在面临对方首次打击的威胁时更倾向于进攻而不是防御来维持战略层面的稳定性。^⑤ 事实上，为了减少对方首先发动核攻击所造成的不安全感，另一方则更倾向于采取先发制人的行动，而这种进攻性战略将破坏国家间的核威慑战略，致使已有全球核安全体系变得不再稳定。

① 黄忠：《智能武器：国际权力斗争的新动力》，《教学与研究》2020年第9期，第78—79页。

② James S. Johnson, “Artificial Intelligence: A Threat to Strategic Stability,” *Strategic Studies Quarterly*, Vol. 14, No. 1, 2020, pp. 22-23.

③ 傅莹：《人工智能对国际关系的影响初析》，《国际政治科学》2019年第1期，第12页。

④ Steven Miller, “A Nuclear World Transformed: The Rise of Multilateral Disorder,” *Journal of the American Academy of Arts & Sciences*, Vol. 149, No. 2, 2020, p. 29.

⑤ Kenneth Payne, “Artificial Intelligence: A Revolution in Strategic Affairs?” *Survival*, Vol. 60, No. 5, 2018, pp. 25-26.

（三）增加危机意外升级的可能性

人工智能技术对核国家彼此间威胁、能力和决策可信度产生的不利影响，将加剧敌对国家间误判、误解及首先发动攻击的风险。一方面，人工智能的复杂性极易引发意外行为和事故，削弱核指挥、控制和通信系统的可靠性，从根本上威胁核国家的信誉并增加意外核冲突的风险；另一方面，核指挥、控制和通信系统存在被黑客网络攻击的风险，在危机时刻使核指挥官失去关键信息和通讯联系，加剧危机升级。兰德公司的研究指出，“如果黑客、信息攻击和数据操纵等行为出现在主导核武器系统的人工智能上面，那么结果将不堪设想。”^①

人工智能军事化应用还可能在危机中创造先发激励机制，模糊防御和进攻之间的界限，增加危机演变为核战争的可能性。在人工智能时代，拥有领先技术的国家更倾向于使用智能化武器首先攻击对方，而相较弱势的国家同样易于发动先发制人的打击，双方在危机中有意或无意发生冲突或核战争的可能性将不断增加。此外，大国的核指挥与控制系统的智能化决策支持程序并非完全可靠。如海上有意或无意的碰撞，依赖人工智能系统可能导致人类作出不成熟或错误的决定和行动，进而影响全球的战略稳定。^②

在危机爆发的情况下，动用或提供人工智能情报、监视及侦察（ISR）系统，也将加剧局势紧张，增加冲突意外升级的可能性。危机期间频繁使用针对敌方核战略设施及核指挥控制系统的人工智能情报收集，极易被敌方误解为是对其核力量进行先发制人攻击的前奏。对方为避免核设施遭意外袭击，就可能会采取核战争行动来获取战略主动权。此外，核国家对人工智能技术增强其核指挥、控制和通信（NC3）系统的过度信任也可带来意外风险。在人工智能军事化应用的背景下，技术在危机中有可能加快战争的速度，减少指挥官的决策时间并带来压力，缺乏深思熟虑可能让他们在核危机期间作出核战争的错误决策。

（四）鼓励军备升级和军备竞赛

^① Edward Geist and Andrew J. Lohn, *How Might Artificial Intelligence Affect the Risk of Nuclear War?* Santa Monica: RAND., 2018, pp. 19-20.

^② Vincent. Boulanin, Lora Saalman and Peter Topychkanov, eds., *Artificial Intelligence, Strategic Stability and Nuclear Risk*, Stockholm: SIPRI, June 2020, p. 4.

当前，人工智能的全球投资越来越被描述为一场军备竞赛。在人工智能时代，一个拥有核武器的国家在某一特定技术上的进步，都会对其他国家产生连锁反应，被视为不稳定的军备竞争。特别是当这项技术的应用范围和变革力量迅速增长，技术演变的速度远远超过国际军备控制规则制定的速度时，这种不稳定将会进一步遭到破坏。相对于核武器技术，人工智能技术在军事领域中的应用缺乏国际普遍接受的管控机制，所引发的军备竞赛也远比美苏冷战时期的核军备竞赛更加复杂多变，为全球战略稳定态势的发展注入了更多不确定性。^①

以军用人工智能的先发制人取代核威慑也会导致军备竞赛的不稳定。自二战结束以来，核大国始终将构建有效的防御系统作为优先任务，在某种程度上维护了全球战略稳定。然而，人工智能引发的军事技术变革使所有核国家的核力量变得愈加脆弱，各国开展军备竞赛的意愿和动机明显加深。如果人工智能技术使核力量变得更容易攻击，在其他条件相同的情况下，拥有核武器的国家需要部署更有能力的报复性武库，以应对日益增长的风险。反作用力技术的快速变化增加了对对手未来能力的不确定性，各国为此需要保留多样化的报复力量，以防范对手技术上的突破。

“军备竞赛的后果是消极的，除了要付出高昂代价，更严重的是破坏稳定，产生反作用，甚至造成灾难”^②，人工智能军备竞赛无疑在验证亨廷顿的这一论断。人工智能军备竞赛所带来的影响主要表现为政府较高水平的开支、不安全的恶性循环威胁以及人工智能开发的无止境三个方面。^③这种高投资率、缺乏透明度、相互猜疑和恐惧以及先发制人的部署动机增加了发生意外冲突的风险。更为严重的是，考虑到人工智能技术所带来的战略优势和红利，世界各主要国家很难凭借自身理性来克制和减缓人工智能军备竞赛的

① 胡高辰：《从核战略态势视角看国际核秩序》，《国际政治科学》2018年第1期，第69页。

② Samuel P. Huntington, “Arms Races: Prerequisites and Results,” *Public Policy*, Vol. 8, No.1, 1958, pp. 41-86.

③ Melanie Sisson, “Multistakeholder Perspectives on the Potential Benefits, Risks, and Governance Options for Military Applications of Artificial Intelligence,” in Brian Finlay, Benjamin Loehrke, and Chris King, eds., *The Militarization of Artificial Intelligence*, New York: Stanley Center for Peace and Security, 2019, p. 4.

步伐，全球更有可能陷入无序和恶性竞争的不稳定状态。

四、人工智能军事化时代如何维护全球战略稳定

针对军用人工智能对现有的全球战略稳定可能产生的冲击和影响，国际社会应聚焦于人工智能军事化过程中相关技术、结构、机制和治理方面建设，通过搭建军用人工智能战略稳定框架，降低人工智能军事化应用过程中对全球战略稳定产生的负面影响，为建立新型全球战略稳定体系创造有利条件。

（一）提升安全可靠可控的技术能力

人工智能军事化应用正处于初始的成长期阶段，技术本身的安全性、可靠性和可控性等很多关键问题还未得到有效解决。固有技术的缺陷不仅表现为人工智能军事系统的适应性不强，容易发生故障、事故和意外的紧急情况，同时也体现在军事系统有限的运行环境感知能力和判断能力。^① 由于人工智能技术的不稳定性极易造成军用人工智能系统数据异常、数据污染和数据泄露等安全风险，人工智能军事系统会经常以意想不到或难以理解的方式运行。“算法黑箱”也将引发人类脱离对人工智能军事系统的控制，呈现出更大的不可预测性。

从技术层面解决人工智能的安全问题，就需要设计使用安全、可靠和可控的军用人工智能系统。首先，任何基于人工智能的军事系统，都应进行可靠的测试和评估，以确定系统的能力和局限性。其次，无论在何种紧急和复杂的情况下都避免使用完全自主运行和决策的人工智能系统，确保军用人工智能始终被人类所控制。再次，提高全球层面人工智能算法的透明性，探寻适用于各国人工智能军事化应用科学统一的可解释标准。

从具体实践来看，世界各国目前根据自身军事需要采取军用人工智能技术开发、部署和使用的不同做法，加大了全球层面应对人工智能技术挑战的难度。鉴于这一现实，国际社会有必要对安全可靠可控的军用人工智能技术进行讨论和谋划。一方面，根据人工智能技术的成熟程度，将人工智能自主

^① Paul Scharre, “Killer Apps: The Real Dangers of an AI Arms Race,” *Foreign Affairs*, Vol. 98, No. 5/6, 2019, p. 135.

武器系统进行统一分类和分级管控，严格禁止将不成熟的人工智能系统投入到实际战场；另一方面，在人工智能军事应用的研发、生产和应用部门间建立不同的技术能力监管层级，根据开发、测试、部署和使用任何特定人工智能军事应用的周期，为全球制定统一的技术监管标准。

（二）夯实战略稳定结构体系

在战略稳定体系的构成要素中，既包括核威慑稳定，也包括人工智能军事化过程中常规武器的威慑稳定，为此需要对既有的全球危机管控体系和国际军备控制体系进行补充和完善。在危机管控方面，国际社会应建立增强战略稳定性的信息共享和提高透明度的信任措施，防止新的敌对行动及管控危机升级。^① 这些信任措施可包括不使用配备核武器的无人作战平台，不使用具有风险标识的情报侦查系统和辅助决策系统，不使用核武器和常规武器首先发动攻击等建立预防危机和缓解危机的重大行动。

在军备控制方面，人工智能自主武器系统、军事决策系统及网络军事行动应纳入新的军备控制体系。一是推进自主武器国际军控讨论，制定规范自主武器系统的标准、指导方针或法规，防止军用人工智能领域军备竞赛走向失控。二是把人工智能决策支持系统归为“战争手段或方法”的管控范围，使这些系统符合法律审查的要求。^② 三是推动搭建网络信息共享和防御黑客网络袭击的平台，建立彼此相互信任的激励措施，通过签订“网络武器”多边军备控制协议，构建基于人工智能的网络军控体系。

当前，美国已将人工智能军事化应用视为大国战略竞争的重要砝码，在一定程度上加剧了全球安全和战略稳定的现实困境。^③ 中俄等国有必要采取切实可行的措施对美国的军用人工智能的发展予以监督和制约。一方面将军用人工智能所涉及的安全问题纳入现有中美、中俄双边战略稳定谈判，顾及

^① Michael C. Horowitz, Lauren Kahn, and Casey Mahoney, “The Future of Military Applications of Artificial Intelligence: A Role for Confidence-Building Measures?” *Orbis*, Vol. 64, No. 4, 2020, pp. 528-543.

^② Giacomo Persi Paoli, Kerstin Vignard, David Danks, and Paul Meyer, *Modernizing Arms Control: Exploring Responses to the Use of AI in Military Decision-making*, United Nations Institute for Disarmament Research, 2020, p. 21

^③ 孙海泳：《美国人工智能军事化的发展态势、风险与应对路径》，《国际论坛》2022年第2期，第37—46页。

有关战略稳定问题的紧迫关切；另一方面，在坦诚交流基础上，就不通过单方面部署非安全、未测试或不可靠的人工智能系统来谋求军事绝对性优势达成共识，确保全球主要大国间的人工智能军事战略力量相对平衡。^①

（三）加强全球多边安全机制建设

根据人工智能军事化应用发展态势，当务之急在于全球层面的多边安全机制建设。针对人工智能技术滥用与现行国际法和习惯法相悖的情形，可在联合国的框架内对原有法律补充和修订，如在《特定常规武器公约》中进一步具体和细化军用人工智能的规范、原则，在《禁止核武器条约》中增加“不使用人工智能自主系统首先攻击核设施”等条目，以更好适应人工智能技术发展的新环境。各国还应健全人工智能武器扩散、转让和出口管控机制，以联合国《武器贸易条约》为基础，商讨和编制人工智能军用技术和产品的多边出口管制清单，防止极端恐怖组织等非国家行为体获得和滥用这些技术。

建构一个全面完整的多边机制框架，是全球军用人工智能迈向规范化的重要路径。面对人工智能技术发展的新趋势，各国需要充分认识和评估人工智能在军事应用过程中可能引发的法律法规、道德标准、伦理规范等方面问题，通过协商谈判、凝聚共识并整合相关法律文书，推动军用人工智能专门国际法律法规的达成。目前，世界各主要国家及国际组织已经就人工智能的军事应用制定了规范性文件。^②从内容来看，各方提出关于军用人工智能的规范和原则存有多个共同点。下一步，有必要在各自关切和管控分歧的基础上，推动制定国际普遍适用的人工智能军事化应用多边安全行为准则。

另外，建立官方战略对话机制和非官方二轨对话机制将有助于缓解人工智能军事化所带来的安全困境。各国军方直接参与解决人工智能和自主系统问题的对话，可作为其本身的一种有价值的沟通机制。^③实际上，“日内瓦国际裁军会议”“慕尼黑安全会议”“香格里拉对话会”等多边交流对话机

^① Imbrie and Kania, *AI Safety, Security, and Stability Among Great Powers: Options, Challenges, and Lessons Learned for Pragmatic Engagement*, Center for Security and Emerging Technology, 2019, pp. 4-17.

^② 李驰江：《人工智能在军事领域的应用及全球治理》，《人民论坛·学术前沿》2021年第10期，第26页。

^③ Michael C. Horowitz and Paul Scharre, *AI and International Stability: Risks and Confidence-Building Measures*, Center for a New American Security, 2021, p. 13.

制，都可用作各国深入探讨人工智能军事化应用的风险挑战及其应对之道的强有力合作平台。中美俄等大国围绕人工智能军事应用议题进行非官方的二轨对话建立更紧密的反馈机制，与政府分享管控人工智能武器风险和国家安全治理路径，也有助于推动实现全球战略稳定新的目标。^①

（四）推动人工智能安全治理合作

鉴于人工智能的军事化应用，全球各方各界已对军用人工智能的安全治理开展广泛讨论，并明确将人工智能安全治理问题纳入各自战略规划，为全球人工智能安全治理提供了潜在基础和必要条件。^② 从全球层面来看，目前所形成的安全治理架构仍呈现出分散性和异质性的特点，很大程度上未能满足人工智能安全治理全球化的发展需求。事实上，只有凝聚人工智能治理的合作共识，推动构建具有普遍意义的人工智能治理框架和人工智能安全治理体系，才能真正确保全球人工智能军事应用的安全和稳定。

全球人工智能安全治理首先要构建多元共治的人工智能安全治理体系，除推动各国政府、国际组织、企业界、学术界等多个利益攸关方的协同参与外，需兼顾发展中国家同西方发达国家的共同参与。全球人工智能安全治理还应把安全监管放在突出位置，通过搭建全球人工智能军事应用多边监管合作平台，构建人工智能全球安全治理的监管框架和监管体系。此外，全球人工智能安全治理的价值体系构建也是国际社会亟待关注的重要问题，各国应在顾及人工智能军事应用理念差异的基础上，尽快制定人工智能军事研发和应用的全球通用价值准则。

值得注意的是，大国在全球安全治理方面始终扮演关键性和引领性的角色。作为全球人工智能技术研究和应用的领先者，中美两国有必要加强在全球人工智能安全治理领域的协调与合作，就如何建立全球层面的人工智能安全治理机制进行讨论，为全球军用人工智能技术治理合作起到示范作用。就中国自身而言，人工智能安全治理是国家关注的重要问题，推动人工智能安

^① 傅莹：《人工智能治理与国际安全治理路径探讨》，《人民论坛》2020 年第 36 期，第 7 页。

^② 李驰江：《人工智能在军事领域的应用及全球治理》，《人民论坛·学术前沿》2021 年第 10 期，第 25—26 页。

全治理的国际合作也是未来关注的重点方向。未来，中国宜继续本着共商、共建、共享的全球治理观和“智能向善”的理念，积极引领全球人工智能安全治理的发展进程，在人工智能领域推动人类命运共同体的构建。

结束语

传统全球战略稳定源于冷战时期美国和苏联之间战略核武器的博弈，并建立在核威慑理论的基础之上。后冷战时期，随着国际战略环境变化及新兴战略技术发展，全球战略稳定范式发生了新的变化。作为第四次工业革命的核心要素，人工智能所引领的技术革命正在延伸至军事领域，增加常规武器和核武器威慑力，并对以核力量为基础的全球战略稳定产生颠覆性影响。人工智能在全球军事化应用领域的迅速扩散及其所带来的负面效应值得我们高度关注。当前，大国间的战略竞争不断加剧，世界主要大国都在顺应时代发展潮流，将人工智能军事化视为在未来战场上取得绝对性军事优势的途径，并竭力将人工智能用于军事目的。在可预见的未来，如果世界军事强国盲目地沿着研发更智能武器的道路前进，那么很可能会把人工智能的军事应用变成通往世界末日的“捷径”。正因为如此，国际社会应聚焦于人工智能军事化过程中的技术、结构、机制和治理等方面建设，为建立起新型全球战略稳定体系创造有利条件。作为全球负责任的新兴科技大国，中国也应加快引领全球战略稳定建构的进程，为推动人工智能时代的全球战略稳定贡献中国智慧和方案。一方面，积极参与人工智能全球军事合作，推动制定具有普遍约束力的人工智能国际安全规范，为国际社会维护全球安全与稳定提供坚实基础；另一方面，积极引领人工智能全球安全治理，推动构建人类核安全命运共同体，为人工智能时代的全球战略稳定提供坚实动力。

[责任编辑：樊文光]