

2021 年第 7 期（总第 25 期）

国际战略与安全研究报告

INTERNATIONAL SECURITY AND STRATEGY STUDIES REPORT

人工智能军事化的挑战及中美合作应对探讨



清华大学战略与安全研究中心

CENTER FOR
INTERNATIONAL SECURITY AND STRATEGY
TSINGHUA UNIVERSITY

人工智能军事化的挑战及中美合作应对探讨

朱荣生^①

随着人工智能技术的快速发展，世界主要大国纷纷制定和发布战略规划，力图抓住智能化浪潮带来的军事变革机遇。有研究认为，美国、中国、俄罗斯、法国、以色列和韩国的军事部门正在加大对人工智能的研发投资，联合企业和科研机构落实研究方案，希望在军事智能化的道路上不落后于人。^②据统计，美国国防部的人工智能投资已从2016财年的约6亿美元增长到2020财年的9.27亿美元，以推动600多个正在进行的人工智能项目。^③而与“人工智能军备竞赛”升温相伴而生的是国际社会对管控军用人工智能安全风险的迫切关注。2021年11月15日，美国国防部国防创新小组发布《负责任的人工智能实践指南》，进而加速美国国防部的人工智能道德原则落实到其商业原型设计和采购工作中。2021年10月21日，北约防长会通过了首份《人工智能战略》，提出“负责任使用人工智能军事应用”的原则。2021年12月13日，中国裁军大使李松在联合国《特定常规武器公约》第六次审议大会上提交了中国《关于规范人工智能军事应用的立场文件》，提出了在智能时代维护国际稳定的治理主张。2021年末紧密出台的监管政策预示着，如何管控人工智能这项颠覆性技术冲击国际稳定将是全球治理下一阶段的重要议题。

① 朱荣生：清华大学战略与安全研究中心助理研究员。本文受到“人工智能的国际安全挑战及其治理研究”（项目号：2020M680492）的资助。

② Frank Slijper, Alice Beck and Daan Kayser, “State of AI: Artificial Intelligence, the Military, and Increasingly Autonomous Weapons,” *Pax*, April 2019.

③ “Defense Budget Overview: United States Department of Defense FY2021 Budget Request,” Office of the Under Secretary of Defense (Comptroller)/Chief Financial Officer, February 2020; Brendan McCord, “Eye on AI,” August 28, 2019, <https://static1.squarespace.com/static/5b75ac0285ede1b470f58ae2/t/5d6aa8edb91b0c0001c7a05f/1567>.

2019年6月以来，清华大学战略与安全研究中心同美国布鲁金斯学会在“人工智能与国际安全治理”框架下进行了五轮二轨国际论坛。尽管研究项目没有把安全困境作为讨论的核心议题，但是资深的军事和外交从业者在交流过程中或多或少表露出对中美卷入一场“非本意”的军事冲突的忧虑。基于我们的研究发现，本文将首先介绍人工智能军事化引发的一般性安全风险，而后列举人工智能时代的安全困境结构要素。在规制军用人工智能国际规范处在摸索阶段的当下，这些发现或许能为推动中美在这一领域的治理合作带来有益启发。

一、人工智能军事化的安全挑战

有战略学者对人工智能的军事效能有很高的期待，认为它如同核武器横空出世一般，注定改变国际安全范式。^④一些军事安全和技术研究者则态度保守，将之视为既有武器的“放大器”，认为它难以兑现人们对它的过高期待。^⑤或许是因为这项正在发展的技术还没有展示出其全部军事效能，人们对它引发的国际安全挑战争论不断。但以下几点趋势得到不少分析人士的认同。

首先，人工智能有可能放大进攻优势，冲击战略稳定。人工智能技术可以大幅度提高侦察数据的收集、目标识别的准确性、目标打击的有效性。例如，利用人工智能技术提高识别核导弹发射井水平，并

^④ Kareem Ayoub and Kenneth Payne, “Strategy in the Age of Artificial Intelligence,” *Journal of Strategic Studies*, Vol. 39, No. 5-6, 2016, pp. 793-819; Kenneth Payne, “Artificial Intelligence: A Revolution in Strategic Affairs?” *Survival*, Vol. 60, No. 5, 2018, pp. 7-32.

^⑤ James Johnson, “Artificial Intelligence and Future Warfare: Implications for International Security,” *Defense & Security Analysis*, Vol. 35, No. 2, 2019, pp. 147-169; Greg Allen and Taniel Chan, “Artificial Intelligence and National Security,” *Cambridge Belfer Center for Science and International Affairs*, July 2017.

使用低成本、高隐蔽性的无人机对重要军事设施实施打击。^⑥有研究悲观地指出，虽然人工智能可以带来意想不到的军事效果，使实施技术强大的一方以奇致胜，但这反而加大了对战略意图的判断难度。^⑦决策者也可能错误地认为，人工智能结合移动和自主传感器平台总能威慑敌方的核报复能力，因而有恃无恐地炫耀武力，并造成危机不断升级。^⑧美国兰德公司认为，人工智能的军事应用会削弱构成核战略理论的主要根据，迫使依赖报复打击的国家不得不面临要么提前发动进攻，要么输掉战争的困境。^⑨

其次，技术扩散引发新的国际安全风险。现有使用“人工智能军备竞赛”的研究既特指国家争相加快军备智能化脚步的竞逐，也指代不断上升的研发投入、人才储备等科技竞争，而中美战略竞争备受瞩目。^⑩以无人机为例，超过30个国家已有或正在开发无人机，至少90个国家和非国家行为体拥有非武装无人机。^⑪与此同时，反无人机产

⑥ Richard Marcum, et al., “Rapid Broad Area Search and Detection of Chinese Surface-to-air Missile Sites Using Deep Convolutional Neural Networks,” *Journal of Applied Remote Sensing*, Vol. 11, No. 4, 2017.

⑦ Paul Scharre, “A Security Perspective: Security Concerns and Possible Arms Control Approaches,” *UNODA Occasional Papers*, No. 30, Nov. 2017, p. 26.

⑧ James Johnson, “Artificial Intelligence and Future Warfare: Implications for International Security,” *Defense & Security Analysis*, Vol. 35, No. 2, 2019, p. 153.

⑨ Jurgen Altmann and Frank Sauer, “Autonomous Weapons and Strategic Stability,” *Survival*, Vol. 59, No. 5, 2017, pp. 121-127; Vincent Boulanin, “The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk. Vol. I Euro-Atlantic Perspectives,” *Stockholm International Peace Research Institute*, 2019.

⑩ Stephen Cave, Seán S. ÓhÉigeartaigh, “An AI Race for Strategic Advantage: Rhetoric and Risks,” *AIES Conference*, Feb. 2-3, 2018; Tim Hwang, Alex Pascal, “Artificial Intelligence Isn’t an Arms Race,” *Brookings Institute*, Dec. 11, 2019; Michael Auslin, “Can the Pentagon Win the AI Arms Race?” *Foreign Affairs*, Oct. 19, 2018; James Johnson, “End of Military-techno Pax Americana? Washington’s Strategic Responses to Chinese AI-enabled Military Technology,” *The Pacific Review*, Vol.13, No. 4, 2019.

⑪ Elisa Catalano Ewers, Lauren Fish, Michael C. Horowitz, Alexandra Sander, and Paul Scharre, “Drone Proliferation: Policy Choices for the Trump Administration,” *Center for a New American Security*, June 2017.

业也急剧增加。据统计，已有33个国家的155家制造商生产制造230多种反无人机系统。^⑫除了军备竞赛风险外，还有政策研究担忧，人工智能被恐怖组织恶用，将之作为报复行动的工具。^⑬例如，“伊斯兰国”（ISIS）曾于2016年在伊拉克北部实施无人机攻击；2018年8月，委内瑞拉总统马杜罗遭到无人机袭击。

再次，人工智能军事化挑战既有交战规则。美国在《特定常规武器条约》官方专家组会议上递交的材料认为，致命性自主武器可以提高军事打击精度，从而减少对平民的伤害。^⑭从这一角度出发，人工智能赋能的武器平台似乎可以更好地践行保护平民的国际人道主义精神。另一方面，人工智能赋能的武器平台可能成为“平民杀手”。谢菲尔德大学的诺尔·夏基（Noel Sharkey）断定，识别技术缺陷使自主武器不能区分合法和非法的攻击目标，更没法判断人类的意图。在复杂的作战环境中，这种缺陷将导致自主武器杀死捡起武器而没有攻击意图的平民。或者，算法偏见可能导致自主武器将平民错误识别为作战人员发动进攻。不论是哪种场景，这都有违国际人道主义法的区分原则。^⑮

最后，人工智能技术存在缺陷，会引发安全风险。尽管民用人工智能技术已经展现出不凡的能力，却并非完美无缺，甚至闹出令人啼

^⑫ Arthur Holland Michel, “Counter-Drone Systems,” *Center for the Study of The Drone at Bard College*, Feb. 2018.

^⑬ Miles Brundage, et al., “Algorithms and Terrorism: The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation,” *UNCCT and UNICRI*, 2018.

^⑭ “Human - Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems,” *United Nations Conference on Certain Conventional Weapons*, August 28, 2018; “Implementing International Humanitarian Law in the Use of Autonomy in Weapon Systems,” *United Nations Conference on Certain Conventional Weapons*, March 28, 2019.

^⑮ Noel Sharkey, “Grounds for Discrimination: Autonomous Robot Weapons,” *RUSI Defence Systems*, Oct. 2008.

笑皆非的情况。例如，谷歌的图像识别软件曾经把黑人识别为“黑猩猩”。^{①⑥}有观点认为，拥有保守文化的军事组织会选择安全和确定性较高的技术而不是先进的人工智能技术，而应用受限则意味着相关安全风险效应的降低。^{①⑦}另一类观点则强调，这不能否认人工智能已经融入到军事领域以及其内在缺陷引发的危机。当前以机器学习带动的人工智能浪潮需要大量数据。基于大数据训练的算法和数据训练集会将偏见无意间带入系统，使人工智能给决策者提供错误的军事建议。更加危险的是利用人工智能赋能网络攻击破坏核武器指挥控制系统。

二、人工智能竞赛加深中美安全困境

安全困境是指一国试图增加自身安全所采取的手段不经意间引发他国的不安全感。^{①⑧}在互信不足和缺少沟通渠道的情况下，一国选择不断增加军备和建立防御措施来维护国家安全，而这会让他国的安全感被大幅降低，并因此而选择扩充军备，最终造成所有人的安全减损。尤其是一国扩张进攻性武器，误判他国意图而实行强硬政策将加深安全困境的程度。在大国加速军事智能化迭代的浪潮中，安全困境理论视角或许可为我们深刻理解维护大国关系稳定提供有益思考。

第一，智能战场压缩决策时间可能令中美军事意外以违背双方意愿的速度升级。2020年7月12日，约翰·艾伦和达雷尔·韦斯特发表

^{①⑥} 《谷歌把黑人标记成大猩猩引争议的解决方式出人意料》，2018年1月18日，<http://tech.sina.com.cn/i/2018-01-18/doc-ifyquptv7683429.shtml>。

^{①⑦} 《谷歌把黑人标记成大猩猩引争议的解决方式出人意料》，2018年1月18日，<http://tech.sina.com.cn/i/2018-01-18/doc-ifyquptv7683429.shtml>。

^{①⑧} John Herz, *International Politics in the Atomic Age*, Columbia University Press, 1959; Robert Jervis, “Cooperation Under the Security Dilemma,” *World Politics*, Vol. 30, No. 2, 1978; Tang Shiping, “The Security Dilemma: A Conceptual Analysis,” *Security Studies*, Vol. 18, No. 3, 2009.

评论文章称，人工智能的军事应用极大地将人类在“观察-定向-决策-行动”（OODA）循环中的决策时间压缩到近乎是即时反应，由此形成的“极速战”迫使竞争对手几乎没有足够的时间对危机作出合理判断。他们认为“极速战”效应会激化中美在南海的紧张对峙中出现难以管控的危机。^{①⑨}美国兰德公司曾以战争推演的方式模拟中美在东亚大量部署无人装备的场景，其推演结果是两国在缺少沟通和高速运转的战场上失去了对危机的掌控力而最终引发了一场局部战争。^{②⑩}这些军事判断和战争推演的逻辑是国家为了获得军事行动中的时间优势，或者确保自身有足够的“快速反应”能力，而越发依赖于技术“操作”弥补人类即时反应的不足，其结果是危机以超脱人类反应的速度螺旋上升。^{②⑪}

第二，美国将人工智能技术部署于危险的进攻性武器平台。曾任美国国防部副部长的罗伯特·沃克（Robert Work）表示，美国不会将致命决策权交予机器，但如果战略竞争对手（尤其是中国和俄罗斯）比美国更愿意把这种权力交给机器，美国将不得不重新审视这一决定。^{②⑫}2019年9月25日，时任美国联合人工智能中心主任杰克·沙纳汉（Jack Shanahan）中将公开表示，美国将确保继续由人类控制核发射决策。^{②⑬}2021年1月13日，美国国防部长阿什利·卡特（Ashton Carter）接受采访问时强调，只有总统而不是机器才能够下达核打击

^{①⑨} John Allen and Darrell West, “Op-ed: Hyperwar Is Coming. America Needs to Bring AI into the Fight to Win — With Caution,” *Brookings Institution*, July 12, 2020.

^{②⑩} Yuna Huh Wong, et al., “Deterrence in the Age of Thinking Machines,” *Rand Corporation*, Jan. 27, 2020.

^{②⑪} Mary Cummings, “Automation Bias in Intelligent Time Critical Decision Support Systems,” *AIAA 1st Intelligent Systems Technical Conference*, 2004, pp. 557–562.

^{②⑫} “David Ignatius and Pentagon’s Robert Work on the Latest Tools in Defense,” *Washington Post*, March 30, 2016, <https://www.washingtonpost.com/>.

^{②⑬} Sydney Freedberg, “No AI For Nuclear Command & Control: JAIC’s Shanahan,” *Defense Breaking*, Sep. 25, 2019, <https://breakingdefense.com/2019/09/no-ai-for-nuclear-command-control-jaics-shanahan/>.

的命令。^{②4} 尽管美国国防部高官释放军事克制信号并展露出对军备竞赛的担忧，但是美军却刻意回避对可搭载核武器的平台进行智能化升级的问题。例如，美国空军将人工智能技术部署于F-35和可搭载核弹的B-2轰炸机以改进他们的攻击目标瞄准系统和忠诚僚机的协同。^{②5}

第三，错误认知让美国误判中国有“颠覆既有国际秩序”的意图。有观点认为，美国将中国的人工智能发展看作是对其全球霸权的挑战，因而对华采取一系列的科技和经贸“规锁”措施。^{②6} 艾尔莎·卡尼亚（Elsa Kania）和罗兰·拉斯凯（Lorand Laskai）认为，美国误判了中国“军民融合”政策的意图，将之误读为中国汲取美国的研发成果，并转化为重塑国际秩序的力量。^{②7} 乔治·华盛顿大学安全与新兴技术中心（CSET）发布报告呼吁政策制定者不应该夸大大数据带来的人工智能发展优势，一些特殊而重要的军事应用场景所产生的数据不多，小数据和先进算法可以为技术带来意想不到的效果，而有效的产业发展政策可以弥补大数据上的劣势。^{②8} 这在一定程度上否定了大数据决定论。而上述研究的潜在政策含义，是美国误判和恶意评估中国的技术发展潜力和政策意图，会将中美推向一场更加严峻的“科技战”，或者陷入赢得了领导力却减少安全的自我挫败中。

^{②4} Nicholas Thompson, “Former DOD Head: The US Needs a New Plan to Beat China on AI,” *Wired*, Jan. 17, 2021, <https://www.wired.com/story/former-dod-head-us-needs-new-plan-to-beat-china-ai/>.

^{②5} Kris Osborn, “Artificial Intelligence Is Going To Make America’s F-35 and B-2 Even Stronger,” *National Interest*, Dec 15, 2019; Warrior Maven, “Air Force brings AI to B-2, F-35 and F-15,” *Fox News*, July 31.

^{②6} Li Zheng, “How to Relieve the Security Dilemma of Sino-US Technology Competition,” *China US Focus*, Oct. 30, 2019; Chris Meserole, “Artificial Intelligence and the Security Dilemma,” *Lawfare*, Nov. 4, 2018.

^{②7} Elsa Kania, Lorand Laskai, “Myths and Realities of China’s Military-Civil Fusion Strategy,” *Center for New America Security*, Jan. 2021.

^{②8} Tim Hwang, Shaping the Terrain of AI Competition, *Center for Security and Emerging Technology*, June 2020; Husanjot Chahal, Helen Toner, and Ilya Rahkovsky, “Small Data’s Big AI Potential,” *Center for Security and Emerging Technology*, Sep. 2021.

第四，美国缺乏适当的政策而盲然采取对华“脱钩”行动。2018年5月和2020年1月，美国国防部和内政部先后以“存在网络安全隐患”为由禁用中国企业大疆生产的无人机。^{②9}由于美国无人机对中国企业的依赖，有专家曾预估“禁飞令”会让美国陷入“无机可用”的尴尬。^{③0}2021年6月1日，五角大楼称，大疆无人机经过测试后没有显示有恶意代码或安全隐患，建议与美国服务机构合作的政府实体和部队继续使用。^{③1}在对华技术“规锁”方面，美国商务部曾公开征集意见，打算将人工智能以技术为单位列入管制清单。这引发了美国产业界的强烈反对。2021年9月，商务部代理副部长杰里米·佩尔特（Jeremy Pelter）在国会听证会上断然拒绝了国会编制这份迟到三年的清单的请求。这些政治“闹剧”表明，如果美国在安全保障上无所作为，就会缺乏安全感；匆忙采取不成熟的政策同样会让美国面临不得不“放弃”的安全困境。

三、中美人工智能全球治理合作路径

结合近期大国密集出台军用人工智能原则的动向以及清华大学战略与安全研究中心的系列二轨对话会成果，以下三点或许能够为政策

^{②9} Brett Forrest and Gordon Lubold, “Air Force Purchase of Chinese Drones Spurs Security Concerns,” *Wall Street Journal*, Nov. 2, 2020, <https://www.wsj.com/articles/air-force-purchase-of-chinese-drones-spurs-security-concerns-11604322017>; Carla Babb, Hong Xie, “U.S. Military Still Buying Chinese-Made Drones Despite Spying Concerns,” *VOA News*, Sep. 17, 2019, <https://www.voanews.com/usa/us-military-still-buying-chinese-made-drones-despite-spying-concerns>.

^{③0} Jack Corrigan, “Interior Approves Chinese Drone Purchases Despite Spying Concerns,” *Nextgov*, July 11, 2019, <https://www.nextgov.com/emerging-tech/2019/07/interior-approves-chinese-drone-purchases-despite-spying-concerns/158361/>.

^{③1} Matt O’ Brien, “Government Use of Chinese Drones in Limbo As Congress Weighs Ban,” *AP News*, June 1, 2021, <https://apnews.com/article/donald-trump-technology-government-and-politics-business-5854cf8b5eccd03f85d5eba2aef10a22>.

制定者管控中美安全困境提供政策参考。

首先，在武器的研发、部署和应用中采纳克制原则，尽快确立军事博弈底线。虽然学者对人工智能武器进行分类、分级管控提出了多种框架，但是总体上都包含禁止能够自我进化、自主攻击人类的机器以及应用于核武器等要素。^{③②}虽然美国国防部高官频繁对外释放不会将人工智能部署于核武器的信号，却将之部署于可搭载核武器的武器平台，并且抛出“持久交手”网络威慑理念将人工智能赋能重点从网络防御转向网络攻击。美国应当承担维护全球稳定的大国责任，在敏感的战略问题上及时修补漏洞，而不是寻求人工智能赋能进攻性武器的替代方案。在《不扩散核武器条约》2020年审议大会上，中国提交的履约报告强调，中国对核力量一直保持绝对安全和绝对可靠的状态，并且不断强化涉核人员的安全意识和责任感。^{③③}或许，这种审慎的自我克制原则和将之纳入到对使用者的教育环节才是划定军事博弈的底线所在。

其次，在官方对话中增加对管控人工智能军事化的交流，减少意图误判的风险。2021年7月26日，美国副国务卿温迪·舍曼（Wendy Sherman）访问中国，提出给中美关系加装“护栏”（guardrail），避免两国“擦枪走火”，中方则明确给出两份“清单”和三条底线的诉求。中美关系似乎进入到重新确立底线的互拉“清单”时代。美国人工智能国家安全委员会的最终报告建议，中美可以建构“中美全面科技对话”的二轨机制，商讨如何应对人工智能技术安全挑战及其军事化对战略稳定的冲击。^{③④}如果能够就人工智能治理建立专门的军备合作对

③② 2021年12月4日，中美专家在清华大学举办的“人工智能与国际安全”专题论坛中提出了多种对军用人工智能管控的框架。

③③ “关于中华人民共和国履行《不扩散核武器条约》情况的国家报告”，中国外交部，2019年4月29日。

③④ “Interim Report,” *National Security Commission on Artificial Intelligence*, Nov. 11 2019; “Final Report,” *National Security Commission on Artificial Intelligence*, March 1, 2021.

话机制，中美可以结合各自的官方文件进行政策交流和对治理原则的正确理解。例如，中国发布的《新一代人工智能治理原则》《新一代人工智能伦理规范》《关于规范人工智能军事应用的立场文件》，美国国防部发布的《人工智能伦理原则》《负责任的人工智能实践指南》，都是两国可交流和讨论的具体内容。

最后，加强中美政策研究界的学术交流，共同探讨军用人工智能原则途径。在清华大学战略与安全研究中心的二轨对话中，中美专家提供了中美学术界对人工智能核心术语定义的文献综述，针对这些内容总体上没有太多不同理解。这说明两国的政策界和学术界有较大的共识空间。尽管双方团队的研究成果对于个别术语的内涵外延存在明显分歧，但是通过拓宽非官方的对话渠道、互换各自翻译后的官方文件并进行政策解读，或许可以为共同研究奠定基础。中美官方层面尚缺少就军用人工智能治理建构专门的沟通机制，此类官方对话能否建立也具有政治不确定性，不过一种可能的方法是中美学者借鉴制定《中美核安全术语》的成功经验，共同探讨《中美人工智能术语》文件。

发表日期：2021年12月14日



扫码关注我们

清华大学战略与安全研究中心

办公地点：清华大学明斋217

联系电话：010-62771388

电子邮箱：ciss@tsinghua.edu.cn