美国军备人工智能化对中美战略稳定的冲击*

丁　伊　　江天骄

【内容摘要】　在人工智能浪潮席卷全球以及中美博弈不断升级的背景下，美国大力推进军备领域的人工智能化进程，并放任其由常规武器领域向核武器领域渗透，以维持自身在全球范围内的军事优势和霸权地位。然而，此举可能对中美战略稳定构成严峻挑战。从危机稳定性方面来看，美国军备人工智能化容易诱发先发制人核打击和危机升级，从而增加中美核战争风险。从军备竞赛稳定性方面来看，美国军备人工智能化可能触发智能化军备竞赛、提升技术管控难度以及增加军备不透明，从而不利于实现世界安全和稳定。为应对这种复杂局势，一方面，中国要大力提升智能化作战能力，确保在必要时能有效捍卫国家安全；另一方面，中国应推动建立人工智能军备的全球治理框架，为国际社会提供行为规范与准则。同时，中美还应努力拓展在人工智能军备领域的合作空间，寻求共同利益与平衡点。

【关键词】　军备　人工智能　中美战略稳定　危机稳定　军备竞赛稳定

【作者简介】　丁伊，上海政法学院上海全球安全治理研究院助理研究员（上海　邮编: 201701）; 江天骄，复旦大学发展研究院副教授（上海　邮编: 200433）

【中图分类号】 D815.1　　**【文献标识码】** A

【文章编号】 1006-1568-(2025)02-0088-22

【DOI 编号】 10.13851/j.cnki.gjzw.202502005

* 本文系上海市哲学社会科学规划课题“美国核武领域的人工智能化研究”（2023EGJ001）的阶段性成果。

随着人工智能技术的不断突破，军备人工智能化已经成为不容忽视的发展趋势。这一重大变革不仅正在重塑世界各国的国防战略和军事部署，还将对大国之间的战略稳定产生深远影响。在国际权力格局发生转变以及中美博弈愈演愈烈的大背景下，美国正在谋求利用人工智能技术维持其绝对军事优势和霸权地位。此举不仅将在技术层面整合美国军事力量的构成，还有可能对中美战略稳定产生负面冲击。

一、文献综述

学术界对于战略稳定的研究主要分为经典战略稳定理论、军事战略稳定理论和复合战略稳定理论三大视角。[①] 就中美战略稳定领域而言，既有研究可按照以上思路划分为以下几个类别。

第一，基于中美核力量的战略稳定分析。该理论视角源于冷战时期旨在稳定美苏对抗强度/烈度的经典核威慑理论，主要探讨如何遏制先发制人核打击，并管控军备竞赛的方式。[②] 从这一理论延伸至中美战略稳定领域，部分学者选择从单一武器系统切入，认为美国导弹防御及其他新型核武器系统可能会削弱中国的核报复能力，进而对中国战略威慑能力构成长期挑战。[③] 还有部分学者致力于从多角度提出影响中美核战略稳定的总体分析框架。[④]

---

① 该分类方法参见：蔡翠红、戴丽婷：《人工智能影响复合战略稳定的作用路径：基于模型的考察》，《国际安全研究》2022 年第 3 期，第 81—87 页。

② Thomas Schelling, *The Strategy of Conflict*, Cambridge: Harvard University Press, 1960; Thomas Schelling, *Arms and Influence*, New Haven: Yale University Press, 1966; Bernard Brodie, *Strategy in the Missile Age*, Santa Monica: Rand Corporation, 1959; Albert Wohlstetter, "The Delicate Balance of Terror: Condensed from Foreign Affairs January, 1959," *Survival*, Vol. 1, No.1, 1959, pp. 8-17; Robert Jervis, "Why Nuclear Superiority doesn't Matter," *Political Science Quarterly*, Vol. 94, No. 4, 1979, pp. 617-634.

③ Fiona S. Cunningham and M. Taylor Fravel, "Assuring Assured Retaliation: China's Nuclear Posture and U.S.-China Strategic Stability," *International Security*, Vol. 40, No. 2, 2015, pp. 7-50; 吴日强、[美]埃尔布里奇·科尔比：《构建中美核武器领域战略稳定》，《现代国际关系》2016 年第 10 期，第 49—50 页; Joshua H. Pollack, "Boost-glide Weapons and US-China Strategic Stability," *Nonproliferation Review*, Vol. 22, No. 2, 2015, pp. 155-164。

④ 李彬、聂宏毅：《中美战略稳定性的考察》，《世界经济与政治》2008 年第 2 期，第 13—14 页；胡高辰：《中美不对称核稳定与美国战略机会主义论析》，《国际安全研究》2021 年第 2 期，第 61—85 页。

基于经典核威慑理论的分析构成了中美战略稳定研究的重要内容。然而，随着新兴技术的不断涌现，这些研究相对忽视新技术手段的缺陷日益凸显。

第二，探究其他新型武器系统对中美战略稳定的影响。此类研究认为，随着军事领域科技水平的不断发展，核因素不再是影响中美战略稳定的唯一军事要素，其他新型常规武器系统，如太空武器[①]、网络武器[②]、高超音速武器[③] 等都有可能冲击中美战略稳定性。然而，此类新型技术虽有可能改变战争面貌，但其颠覆能力相对有限，造成的政治与心理影响力尚不足以与核武器相提并论，更难以形成类似核武器的威慑平衡。[④] 因此仍不能从根本上撼动核武器在全球战略稳定中的决定性作用。目前，相当一部分学者仍然倾向于通过核威慑等传统路径来考察其对战略稳定的影响。[⑤]

第三，基于中美复合战略稳定视角的研究。此类研究认为，军事因素对中美战略稳定的影响并不具有唯一性和排他性。除核威慑等军事力量外，中美之间的经济相互依赖、人文交流、政治互信、危机管控与国际环境等因素理应被纳入，以考察整体环境的稳定性。[⑥] 复合战略稳定理论虽然范围宽泛，但其理论框架相对松散，且该领域几乎所有研究都将核因素纳入战略稳定分析框架，这反映出核领域仍然是确保中美战略稳定的重中之重的客观事实。

---

① 何奇松：《中国太空崛起与中美太空关系》，《美国问题研究》2016 年第 2 期，第 37—57 页。

② 鲁传颖：《网络空间大国关系演进与战略稳定机制构建》，《国外社会科学》2020 年第 2 期，第 96—105 页；江天骄：《中美网络空间博弈与战略稳定》，《信息安全与通信保密》2020 年第 9 期，第 11—17 页。

③ Dean Wilkening, “Hypersonic Weapons and Strategic Stability,” *Survival*, Vol. 61, No. 5, 2019, pp. 129-148.

④ Martin C. Libicki, *Cyberdeterrence and Cyberwar*, Santa Monica: Rand Corporation, 2009, pp. xv-xvi, https://www.rand.org/content/dam/rand/pubs/monographs/2009/RAND_MG877.pdf.

⑤ 徐能武、黄长云：《太空威慑：美国战略威慑体系调整与全球战略稳定性》，《外交评论》2014 年第 5 期，第 62—84 页；吴挺：《从中美战略稳定性看太空武器化问题》，复旦大学硕士论文，2013 年；原玥：《外层空间军备控制研究：跨域威慑与战略稳定》，外交学院博士论文，2023 年；罗曦：《美国构建全域制胜型战略威慑体系与中美战略稳定性》，《外交评论》2018 年第 3 期，第 37—62 页。

⑥ 鹿音：《中美战略稳定关系的演进》，《当代美国评论》2017 年第 2 期，第 20—38 页；王政达：《中美复合战略稳定关系：建构依据、基本框架与发展趋势》，《国际安全研究》2019 年第 5 期，第 79—107 页。达巍：《中美关系与中美中长期战略稳定框架》，《国际关系研究》2016 年第 1 期，第 14—17 页。[美]托马斯·芬加、樊吉社：《中美关系中的战略稳定问题》，《外交评论》2014 年第 1 期，第 43—55 页。

由此可见，学术界在中美战略稳定领域已经取得了较为丰硕的成果。随着理论的不断丰富，一方面，核因素作为战略稳定核心的特质始终未变。因此，对中美战略稳定的研究仍然需要以核武器为抓手，并结合技术发展与国际格局的动态变化与时俱进。另一方面，人工智能技术向包括核武器在内的各个军事领域不断渗透，并深刻影响军力状况和国际权力格局，成为影响战略稳定的新兴不确定因素。战略稳定研究需要将该因素纳入考量范畴，并对既有理论进行丰富和完善。有鉴于此，本文致力于在经典核威慑理论的基础上，充分挖掘人工智能这一新兴技术变量对核武器等军事系统的赋能情况，以及其对中美战略稳定的影响，从而为后续研究开启新的思路。

二、美国军备人工智能化的发展

随着人工智能技术的迅猛发展，其在军事领域的应用潜力愈发显著。在中美博弈的大背景下，美国不断扩大人工智能军备的应用范围，以维护其军事优势与霸权地位。目前，美国的军备人工智能化已出现由常规领域向核武器领域渗透的趋势，极有可能对中美战略稳定产生负面影响。

（一）人工智能技术推动军备领域变革

人工智能是用于模拟、延伸和扩展人类智能的理论、方法、技术及应用系统的一门新的技术科学。[①] 人工智能技术对军备的赋能主要体现在以下三个方面。

首先，战场态势的数字化有利于提升态势感知水平。人工智能可通过大数据分析技术处理多源异构数据，促进目标识别由发现向透视质变。此类技术不仅覆盖物理空间，更能整合网络等虚拟域信息，形成全域战场透明化态势，构建实时的数字化战场地图，从而增强情报分析能力。

其次，人机耦合模式有利于提高决策效率。借助深度学习算法，人工智能可以通过学习海量的历史数据来精准识别特定的模式并预测敌方行为，或提出可能的行动方案辅助人类在复杂战场环境中进行决策。此外，“人工智

① 张智海主编：《制造智能技术基础》，清华大学出版社 2022 年版，第 2 页。

能推演+人类决断”的协同决策模式将重新定义人机分工界面，使推演耗时从“天”压缩到“秒”，帮助人类决策者过滤低价值信息并专注于核心矛盾，从而极大缩短决策时间。

再次，人工智能将显著赋能武器系统。在人工智能加持下，自主武器系统可通过数字孪生技术进行虚拟测试，从而增进攻防效能，帮助其在战场环境下实现规避威胁和精准打击的有效结合。此外，人工智能技术还可帮助其快速整合数字仿真、物理实体和实时数据流，重构杀伤链运行逻辑，将传统线性流程升级为“多线程并进+动态调整”的网状结构，从而大幅压缩杀伤链周期。

由此可见，人工智能的军事化应用不仅能够大幅提升武器的智能化水平，还使得军事杀伤链加速闭合。随着人工智能在战争体系中的广泛应用，作战流程将被重新塑造，军事杀伤链将提速增效，感知快、决策快、行动快等将成为未来智能化战争制胜的重要砝码。[①]

（二）美国人工智能军备应用日益广泛

人工智能技术已被广泛应用于多个军事领域。在作战场景中，指挥官往往需要充分收集环境中的所有相关信息和数据，快速形成对当前形势的理解和判断，制定最佳行动方案并予以实施。这就要求其遵循观察（observe）、判断（orient）、决策（decide）和行动（act）的作战流程，即 OODA 循环。以下将基于 OODA 循环，对美军的人工智能应用领域进行分析和梳理。

第一，观察环节，主要涉及情报、监视与侦察能力。人工智能被应用于快速处理和分析从卫星、无人机、地面传感器、通信拦截以及公开信息等渠道获取的大量图像与数据。2017 年 4 月，美国国防部成立“算法战跨职能小组”（Algorithmic Warfare Cross-Functional Team，亦称 Project Maven）。该项目旨在通过自动分析和解读大量视频数据，提高情报收集和目标识别效率，标志着美军智能化建设走向“快进”模式。[②] 美国密苏里大学研究团队则通过训练深度学习算法从高精度卫星图像中检测、识别中国疑似防空导弹

① 吴明曦：《智能化战争——AI 军事畅想》，国防工业出版社 2020 年版，第 68 页。

② 《算法战：牵引美军人工智能军事化应用》，《解放军报》2017 年 11 月 23 日，第 11 版。

阵地。[①] 这表明人工智能在军事情报领域具有广泛的应用前景。

第二，判断和决策环节，主要涉及态势感知能力和辅助决策。2022 年，美国国防部推出“联合全域指挥与控制”（Joint All-Domain Command and Control，简称 JADC2）概念，旨在将其所有传感器和作战单元实时连接，使各军种内部、不同军种间、美军与盟军间，在陆、海、空、天、网等各个作战域都能实现无缝衔接和协调一致。[②] 在此基础上，人工智能将凭借先进的数据处理和通信技术，提供实时的态势感知和智能分析，从而辅助指挥官作出快速而准确的决策。

第三，行动环节，主要涉及自主武器系统的应用。自主武器系统利用人工智能技术进行目标识别和打击，能够在减少人员伤亡的同时，提高作战效率、响应速度以及打击精度。目前，美军已推出且仍在研制多种自主武器系统，包括无人机作战平台、智能化远程反舰导弹、无人地面作战平台、水面和水下无人系统以及电子战系统等，[③] 该系统在相当程度上已具备无人操控条件下自主执行任务的能力。

（三）美国人工智能军备向核武器领域渗透

由于核武器的大规模杀伤性及其造成的人道主义灾难，人工智能在核武器领域的渗透往往更具争议性。尽管在美国国会和政府中都传出了禁止人工智能操控核武器的声音，[④] 但人工智能技术还是从常规武器领域不断渗透到美国涉核军事行动的各个环节中（见图 1）。

---

① Kyle Mizokami, “This AI Hunts Chinese Missiles Sites,” *Popular Mechanics*, November 22, 2017, https://www.popularmechanics.com/military/research/a13865168/this-ai-hunts-chinese-missiles-sites/.

② 李学华：《美军制定“联合全域指挥与控制”细则》，《中国国防报》2022 年 3 月 30 日，第 4 版。

③ 张梦淅等：《世界国防领域前沿技术发展及应用概览》，中国宇航出版社 2023 年版，第 195—204 页。

④ Robert Hart, “Don’t Let AI Control Your Nukes, U.S. Official Urges China and Russia,” *Forbes*, May 2, 2024, https://www.forbes.com/sites/roberthart/2024/05/02/dont-let-ai-control-your-nukes-us-official-urges-china-and-russia/; *Block Nuclear Launch by Autonomous Artificial Intelligence Act of 2023*, 118th U.S. Congress, May 1, 2023, https://www.congress.gov/bill/118th-congress/senate-bill/1394?s=1&r=2&q=%7B%22search%22%3A%22senate+1394%22%7D.

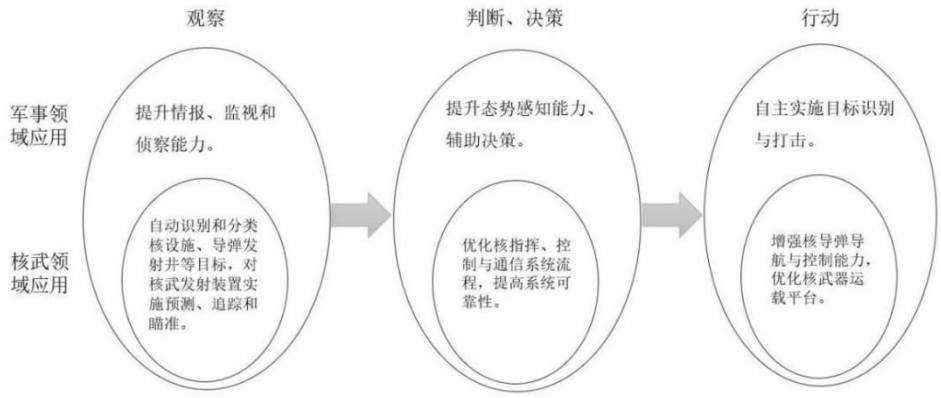

图 1　人工智能在美国军备中的应用

资料来源：作者自制。

第一，在观察环节，美军涉核情报、监视与侦察领域可能将得到人工智能技术的加持。美国可利用人工智能技术处理数据、识别信号，从而自动识别和分类核设施、导弹发射井和其他相关目标，并利用机器学习算法在海量数据中识别出可能存在的异常活动或变化，为核武器发射平台的预测、追踪及瞄准提供有力的情报支撑。例如，美国利用 RQ-170 隐形无人机等从空中秘密投放传感器设备，再使用情报融合系统实时处理和分析数据，生成机动导弹阵地和巡逻模式图像，并不断实时更新，以此威胁对手的第二次核打击能力。[①] 另有报道称，美国军方正在利用人工智能预测核导弹的发射。此外，还有多个人工智能机密项目正在进行，以保护美国免受核导弹袭击。[②]

第二，在判断与决策环节，人工智能技术对美军核指挥、控制与通信系统的渗透在不断加深。该系统是涉核决策的核心环节，人工智能技术能够帮助其优化决策流程并提高系统的可靠性，这有助于决策者在紧急情况下做出最优选择。当前，美国等国家的核指挥、控制与通信系统越来越依赖于专家

① Austin Long and Brendan Rittenhouse Green, “Stalking the Secure Second Strike: Intelligence, Counterforce, and Nuclear Strategy,” *Journal of Strategic Studies*, Vol. 38, Nos. 1-2, 2015, p. 64.

② Phil Stewart, “Deep in the Pentagon, A Secret AI Program to Find Hidden Nuclear Missiles,” Reuters, June 6, 2018, https://www.reuters.com/article/idUSKCN1J114J/.

系统和机器学习算法，以增强信息流、态势感知和网络安全。随着该趋势的加强，人工智能军备竞赛可能会降低战略稳定性。[①] 美国战略司令部司令海滕（John Hyten）认为，人工智能将在下一代核指挥与控制系统架构中发挥重要作用。[②] 美国国防部已多次实施“全球信息优势实验”（Global Information Dominance Experiment），即利用人工智能技术进行模拟的数据整合和实时分析，从而改进数据共享和决策流程，并为美国实施“联合全域指挥与控制”提供技术支持与实践经验。该实验涉及针对可能搭载核弹头的潜艇与导弹的动态预测及辅助决策。[③] 此外，有报道称，美国军方已升级作为关键核指挥、控制与通信系统之一的“战略自动指挥和控制系统”（Strategic Automated Command and Control System）。该系统负责向美国空军各作战部队传递核与常规紧急行动指令及目标数据，目前已集成现代化数据存储与内存技术，以大幅提升战场决策者对关键数据的检索及验证效率。[④] 这些举措旨在确保美军在未来可能发生的核冲突中拥有信息优势和决策优势。

第三，在行动环节，美军可能正在着手对核武器及其运载系统进行人工智能化升级。人工智能技术有助于增强核导弹的导航与控制能力、优化核武器运载平台，从而提升核运载系统的整体性能、打击精度以及在对抗中的生存能力。据报道，美国已为其核弹头设计了新型组件以帮助寻找正确的代码和环境信号来解锁系统，能够在更精准的时机引爆这些弹头，从而增强破坏力。[⑤] 此外，美国还着手对具有携带核弹头作战能力的战斗机和轰炸机等运

① Mark Fitzpatrick, “Artificial Intelligence and Nuclear Command and Control,” *Survival,* Vol. 61, No. 3, 2019, p. 82.

② Colin Clark, “STRATCOM’s Hyten on B-21, Columbia Class, NC3,” Breaking Defense, April 16, 2018, https://breakingdefense.com/2018/04/stratcoms-hyten-on-b-21-colombia-class-nc3/.

③ Michael Evans, “Pentagon Uses AI to Predict Enemy Moves ‘Days in Advance,’ ” *Times*, August 3, 2021, https://www.thetimes.co.uk/travel/destinations/north-america-travel/us/pentagon-uses-ai-to-predict-enemy-moves-days-in-advance-bql5q5s9p.

④ Airforce Technology, “ITT Exelis to Upgrade USAF’s Strategic Automated Command Control System,” Airforce Technology, May 8, 2013, https://www.airforce-technology.com/news/newsitt-exelis-to-upgrade-usafs-strategic-automated-command-control-system/?cf-view; Justin Oakes, “New Construction Begins at Offutt, SACCS to Receive New Home,” Stratcom, March 5, 2020,https://www.stratcom.mil/Media/News/News-Article-View/Article/2104070/new-construction-begins-at-offutt-saccs-to-receive-new-home/.

⑤ R. Jeffrey Smith, “Sensors Add to Accuracy and Power of U.S. Nuclear Weapons but May Create New Security Perils,” *Washington Post*, October 29, 2021, https://www.washingtonpost.

载系统实施人工智能升级。2024 年，美国空军部长肯德尔（Frank Kendall）乘坐了由人工智能控制的 F-16 战机改装机型，进行模拟空战演练，并表示人工智能有能力决定是否在战争中发射武器。[①] 即将投入现役的新一代隐形轰炸机 B-21“突袭者”（Raider）也将在智能化领域取得突破：一是配备最先进的航空电子设备和电子战系统，可通过提高态势感知、通信和防御敌方威胁等措施，提升其在对抗环境中的生存力和作战效能；[②] 二是具备一定的自主飞行能力。B-21 轰炸机计划生产飞行员驾驶和无人驾驶两种机型。[③] 由于其可装载核弹头，不排除其在无人状态下对目标实施核打击的可能性。此外，美军 F-35 战斗机等机型都将进行智能化升级，使其在响应速度、态势感知、辅助决策和协同能力等方面大幅提升。[④]

三、美国军备人工智能化对中美战略稳定的影响

美国积极推进人工智能技术在军事领域乃至核武器领域的应用，可能深刻影响中美战略稳定性，同时为中美博弈带来更多复杂性与不确定性。经典核威慑理论认为，战略稳定性包含危机稳定性（crisis stability）与军备竞赛稳定性（arms race stability）。[⑤] 危机稳定性取决于行为体实施先发制人核

---

com/national-security/us-nuclear-weapons-electronic-sensors-accuracy/2021/10/28/79533ff0-34cc-11ec-9bc4-86107e7b0ab1_story.html; Michael Baker, “With Redesigned ‘Brains,’ W88 Nuclear Warhead Reaches Milestone,” Sandia National Laboratories, August 13, 2021, https://www.sandia.gov/labnews/2021/08/13/with-redesigned-brains-w88-nuclear-warhead-reaches-milestone/.

① Tara Copp, “An AI-Controlled Fighter Jet Took the Air Force Leader for a Historic Ride. What that Means for War,” Associated Press, May 4, 2024, https://apnews.com/article/artificial-intelligence-fighter-jets-air-force-6a1100c96a73ca9b7f41cbd6a2753fda.

② Defense News Aerospace, “US Northrop Grumman's B-21 Raider Stealth Bomber Achieves Milestone in Flight Testing,” Army Recognition Group, May 25, 2024, https://armyrecognition.com/news/aerospace-news/2024/breaking-news-us-northrop-grummans-b-21-raider-stealth-bomber-achieves-milestone-in-flight-testing-2.

③ Harrison Kass, “The B-21 Raider Bomber Question Everyone Keeps Asking,” *National Interest*, December 20, 2023, https://nationalinterest.org/blog/buzz/b-21-raider-bomber-question-everyone-keeps-asking-208064.

④ Volreka F. Senatus, “Pentagon Integrates New AI into F-35 to Fly & Attack into 2070,” Warrior Maven, December 13, 2023, https://warriormaven.com/air/pentagon-integrates-new-ai-into-f-35-to-fly-attack-into-2070.

⑤ Thomas Schelling, *The Strategy of Conflict*, Cambridge: Harvard University Press, 1980,

打击的动机是否强烈。如果先发制人能够取得明显优势，则其中一方首先发动核打击的意愿较高，即危机稳定性较低；反之，则危机稳定性较高。[①] 军备竞赛稳定性取决于扩充核武库的动机是否强烈，即一方扩充军备的某个行为是否会引发对手的跟进并导致军备竞赛。如果该行为容易引起对手扩充军备，则军备竞赛稳定性较低；反之，则军备竞赛稳定性较高。[②] 在人工智能等新技术要素的加持下，中美之间的危机稳定性与军备竞赛稳定性亦将随之发生新的变化。

（一）美国军备人工智能化对中美之间危机稳定性的影响

美国不遗余力推进军备人工智能化可能诱发先发制人核打击，并导致危机的核升级，从而不利于中美之间危机稳定性的构建。

第一，诱发先发制人核打击。人工智能技术在军事领域的运用可能在进攻与防御两个维度上破坏中美攻防平衡状态，从而加强美国实施先发制人核打击的动机。

从进攻角度看，一方面，人工智能技术有助于核武器及其运载平台在决策速率、制导、续航和智能化等方面实现大幅提升，从而为突破对手的防御系统提供了新利器。如前所述，美国多款战斗机与战略轰炸机已实施人工智能升级以提高作战效能。此外，2021 年，美军 MQ-25A 舰载无人加油机采用先进的自动化和人工智能技术提高其自主飞行和操作能力，已成功为 F/A-18F 战斗机等可能搭载核弹头作战的机型实施空中加油，从而提升持续作战能力。[③] 另一方面，人工智能技术还可能用于削弱一方的第二次核打击能力，从而巩固另一方的进攻优势。比如，中国奉行不首先使用核武器政策，坚持自卫防御核战略。然而，人工智能强化的情报、监视与侦察能力或将导致中国的核导弹发射基地、弹道导弹核潜艇等设施更容易被发现、定位和摧毁，由此可能增加第二次核打击力量的脆弱性，并削弱中国的核威慑能力。

---

pp. 205-254; Thomas Schelling and Morton Halperin, *Strategy and Arms Control*, Mansfield Centre: Martino Publishing, 2014, pp. 32-42.

① 李彬：《军备控制理论与分析》，国防工业出版社 2006 年版，第 79 页。

② 李彬：《军备控制理论与分析》，第 83 页。

③ Sam Lagrone, “MQ-25A Unmanned Aerial Tanker Refuels Super Hornet in Successful First Test,” U.S. Naval Institute, June 7, 2021, https://news.usni.org/2021/06/07/mq-25a-unmanned-aerial-tanker-refuels-f-a-18-hornet-in-successful-first-test.

从防御角度看，人工智能技术还有助于强化战略防御体系。美国国防部导弹防御局（Missile Defense Agency）正在将机器学习与人工智能等先进技术应用于导弹防御，以强化系统测试与评估、检测追踪和分辨目标、系统指挥与控制以及打击目标等功能。[①] 在实际应用方面，美国北方司令部（NORTHCOM）和北美防空司令部（NORAD）已引入“探路者”（Pathfinder）系统，利用人工智能和机器学习技术分析北方预警系统中的海量传感器数据，从而提升两司令部情报处理和防御能力。目前，该系统已经展示出有效检测和跟踪小型无人机的能力。未来，北方预警系统将继续推进现代化升级，利用人工智能技术提升对轰炸机、巡航导弹等目标的探测能力。[②]

综上所述，人工智能对于美方战略打击武器的赋能、对中方第二次核打击力量的削弱以及对美方防御力量的强化可能会促使攻防平衡局势向有利于美方的方向转变。而在一方进攻力量占据上风的情况下，核战争爆发的可能性亦会有所上升，从而削弱危机稳定性（见图 2）。[③]

图 2　人工智能应用可能诱发先发制人核打击的作用机制

资料来源：作者自制。

① C. Todd Lopez, “Vice Admiral Discusses Potential of AI in Missile Defense Testing, Operations,” U.S. Department of Defense, August 12, 2021, https://www.defense.gov/News/News-Stories/Article/Article/2730215/vice-admiral-discusses-potential-of-ai-in-missile-defense-testing-operations/.

② Frank Wolfe, “North Warning System Modernization may Include Detection of Bombers, Cruise Missiles and Small UAS,” *Defense Daily*, August 17, 2021, https://www.defensedaily.com/north-warning-system-modernization-may-include-detection-of-bombers-cruise-missiles-and-small-uas/air-force/.

③ Stephen Van Evera, “Offense, Defense, and the Causes of War,” *International Security*, Vol. 22, No. 4, 1998, pp. 5-43.

第二，导致危机升级。随着人工智能技术的不断发展，涉核指挥、控制和通信系统对人工智能辅助决策的依赖程度日益加深。尽管美国官方明确表示，使用人工智能的涉核决策系统将确保“人在回路”，[①] 然而，人类的最终决策将不可避免地依赖人工智能的判断，原因如下（见图 3）。

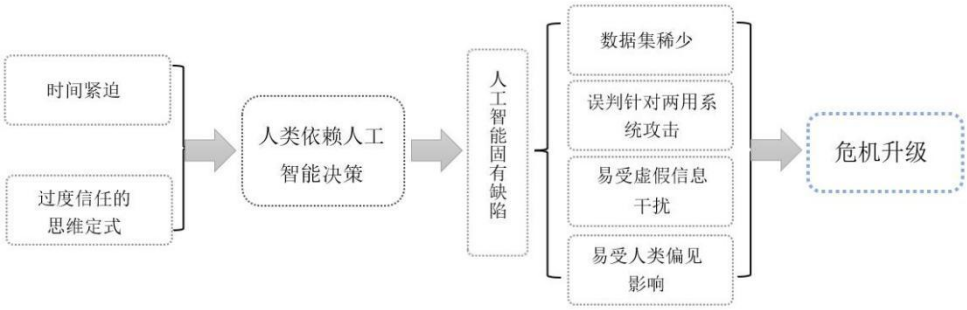

图 3　人工智能应用可能导致危机升级的作用机制

资料来源：作者自制。

首先，时间的紧迫性。现代战场环境日益复杂，作战效能成倍增长，战争节奏显著加快，从而对指挥官的决策效率提出了更高的要求。卡内基梅隆大学团队利用太空资产、现有网络以及人工智能等手段将决策时间从大约 20 分钟缩短到 20 秒。美国军方正在与其保持密切合作并支持其研发。[②] 在重大军事变革之下，指挥官可能会选择向人工智能让渡部分决策权力以适应形势的瞬息万变，主要有以下原因：其一，危机期间的高压环境会对指挥官造成一定的心理压力和认知障碍，尤其是涉核情境通常要求指挥官在极短时间内作出关键决策，否则极有可能面临灾难性后果，这对于人类而言近乎极

① “人在回路”（human in the loop）意味着任何关于发射或使用核武器的决定必须经过人类的确认，而不是完全由自动化系统独立完成，即确保人类对核武器的绝对控制权。参见 *Nuclear Posture Review*, U.S. Department of Defense, October 2022, p. 13, https://media.defense.gov/2022/Oct/27/2003103845/-1/-1/1/2022-NATIONAL-DEFENSE-STRATEGY-NPR-MDR.PDF.

② Lee Hudson, “War Footing,” *Aviation Week & Space Technology*, October 12-25, 2020, p.51.

限考验；其二，人工智能系统能够为实时决策提供必要支持。核危机期间涉及的信息量十分庞杂，而人工智能系统能够相对冷静、高效地整合和处理海量数据，为指挥官提供更加全面的态势感知、情景预测和决策建议。

其次，人类在人机互动的过程中易形成过度信任机器的思维定式。从社会层面来看，人类通常更加厌恶遭到同类的背叛，而对机器背叛的容忍度较高。此外，与机器打交道并不像与人的交往那样可能影响未来的预期。从人机互动层面来看，人类在操作机器的过程中往往存在以下误区：其一，技术的普遍适用使人类默认除数字化之外别无选择；其二，人类倾向于不加批判地接受系统的输出结果而忽视定期检查；其三，人类倾向于假定旧系统中的安全检查和监控习惯已经延续到新系统中，然而事实往往并非如此；其四，决策者对系统潜藏的风险认识不足。[①] 事实上，人类对于机器的过度信任在军事行动中已初现端倪。例如，以色列在对加沙的轰炸中使用“薰衣草”（Lavender）人工智能辅助系统对疑似目标进行分析和瞄准。该系统的操作员表示，他们对于人工智能的决策除了批准之外几乎没有任何意义，基本将所有人工智能输出视为“人类的决定”。[②]

在涉核决策中，虽然人们要求确保核武器的最终发射权由人类掌握，但在实践中，向人工智能让渡部分指挥与决策权或将成为战略竞争领域的新常态。然而，由于人工智能系统的固有缺陷，这种情形可能引发危机的升级乃至失控，从而加剧核战争爆发的风险。

首先，用于涉核决策的人工智能系统训练数据集十分有限，这可能使机器模型无法充分学习和捕捉数据中的模式和规律，导致模型的准确性和泛化能力下降，从而使系统对紧张局势产生误判。其次，由于核常两用设施本身的模糊性质，人工智能未必能够全面理解复杂的战略和政治情境、区分常规攻击或针对核设施的意图，从而显著增加核升级的可能性。再次，电子战的发展以及虚假信息的广泛传播对人工智能决策产生干扰。敌方可通过伪造雷

① Patricia L. Hardré, “When, How, and Why Do We Trust Technology Too Much?” in Sharon Y. Tettegah and Dorothy L. Espelage, eds, *Emotions, Technology, and Behaviors*, Cambridge: Academic Press, 2015, pp. 85-106.

② Yuval Abraham, “‘Lavender’: The AI Machine Directing Israel’s Bombing Spree in Gaza,” *+972 Magazine*, April 3, 2024, https://www.972mag.com/lavender-ai-israeli-army-gaza/.

达信号、干扰通信指令和误导传感器数据等方式向人工智能系统传递虚假信息，以干扰和欺骗其判断和决策，甚至使其发动不必要的核攻击。最后，人工智能容易受到人类偏见的影响，如特定群体的数据被过度代表或被忽略，系统开发者固有的经验和认知等都会对算法的选择和参数设定造成影响，人工智能则将学习并继承这些偏见。在中美竞争的背景下，较高的冲突性和对抗性会成为美国人工智能系统开发时预设的偏见，一旦接受此设定，该系统输出攻击性指令的可能性也会大幅提升。

（二）美国军备人工智能化对中美军备竞赛稳定性的影响

美国军备人工智能化的快速推进可能触发智能化军备竞赛、提升技术管控难度，并增加美方的军备不透明度，不利于中美军备竞赛稳定性的构建。

第一，触发中方反制。美国军备力量的现代化升级旨在巩固其军事优势，并削弱中国在周边地区的作战能力。当中国认为现有军事力量不足以应对军事变革带来的新安全威胁时，就将被迫采取措施来弥补军事力量短板，维护国家安全。

首先，针对美国对我第二次打击能力的威胁，中国提出了打造强大战略威慑力量体系的发展要求，着力提升核力量发展水平。一是加强可能运载核武器的平台的机动性与突防能力，确保其在战争中具有较高的生存能力。如中国研发的新型“东风-41”洲际导弹可携带多枚分导式核弹头，具有较强的突防能力和机动性能，并有望实现实战部署。[①] “东风-17”高超音速导弹具备全天候、无依托、强突防的特点，可对中近程目标实施精确打击。[②] 二是提升第二次核打击力量的隐蔽性以确保其战略威慑能力。

其次，发展新域新质作战能力以削弱对手涉核领域的监视和侦察能力、指挥控制与通信能力以及核打击能力。新域新质作战力量正在深刻塑造战争理念与战场态势，成为大国竞争的新兴赛道和决胜疆场。党的二十大报告强调，“打造强大战略威慑力量体系，增加新域新质作战力量比重，加快无人

① 《央视罕见披露中国洲际弹道导弹东风-41 部分技术超美俄　试射无失败记录》，央视网，2017 年 11 月 27 日，http://m.news.cctv.com/2017/11/27/ARTI9VUMsHTzkFJWzqMH8zM1171127.shtml。

② 《东风-17 常规导弹方队：使命必达的精确打击尖刀》，新华网，2019 年 10 月 1 日，http://www.xinhuanet.com/politics/70zn/2019-10/01/c_1125063244.htm。

智能作战力量发展，统筹网络信息体系建设运用。”[①] 公开资料显示，目前中国已在反潜作战能力[②]、无人机与巡飞弹作战能力[③]、电子战能力[④]、定向能武器开发[⑤] 等方面取得突破，以期对美方实施强有力的反制。

美国利用人工智能技术来谋求军事优势与霸权地位，增加了中方对其军备效能评估的诸多不确定性，引发了中方的安全忧虑，因此不得不针对性地强化自身军事力量以抵消美方施加的安全压力。在中美博弈的背景下，美国的军备升级可能进一步加剧螺旋式上升的“军备竞赛”以及“安全困境”。

第二，政企并进模式增加技术管控难度。军备人工智能化发展呈现出政府与企业并进的特点。历史上，包括核武器在内的许多重大军事技术革新在初始阶段都由政府或依托政府的机构掌握，这使得国家能够有效地约束和管控其在军事应用和产业领域的扩散。然而，在本轮军备人工智能化浪潮中，人工智能技术的源头不再被政府及其主导机构垄断，这种情形可能加大技术管控难度，不利于实现军备控制。

首先，多元化主体增加协调难度。目前，相当一部分国家的军事部门依赖于从商业机构采购的人工智能产品。美国军方与诸多开展互联网、人工智能以及数据分析业务的企业保持着密切合作。[⑥] 在本轮人工智能化变革中，

---

① 习近平：《高举中国特色社会主义伟大旗帜　为全面建设社会主义现代化国家而团结奋斗——在中国共产党第二十次全国代表大会上的报告》，人民出版社 2022 年版，第 56 页。

② 邓彬、李韬、汤斌等：《基于太赫兹雷达的声致海面微动信号检测》，《雷达学报》2023 年第 4 期，第 817 页；Stephen Chen, “Chinese Scientists Look to 6G to Hunt Submarines, Testing Device Small Enough to Fit on Drone,” *South China Morning Post*, August 29, 2023, https://www.scmp.com/news/china/science/article/3232682/chinese-scientists-look-6g-hunt-submarines-testing-device-small-enough-fit-drone。

③ 《“彩虹”无人机总师谈未来无人机战场应用》，环球网，2022 年 11 月 11 日，https://mil.huanqiu.com/article/4AQ8QzaOt7e。

④《〈砺剑〉20240118 奋进深蓝 南昌舰》，央视网，2024 年 1 月 18 日，https://tv.cctv.com/2024/01/18/VIDE3cLeKKXMNO1jrNt9bbA6240118.shtml。

⑤ 《“光箭”“天盾”系列激光安防装备新品亮相珠海航展》，人民网，2024 年 11 月 13 日，http://military.people.com.cn/n1/2024/1113/c1011-40360205.html；高明辉、郑玉权、王志宏：《天基激光武器系统的发展》，《中国光学》2013 年第 6 期，第 810—817 页；《臧继辉委员：简述“反卫星技术”》，中国政协网，2021 年 6 月 24 日，http://www.cppcc.gov.cn/zxww/2021/06/24/ARTI1624501659719410.shtml。

⑥ Drew Harwell, “Google to Drop Pentagon AI Contract after Employee Objections to the ‘Business of War,’ ” *Washington Post*, June 1, 2018, https://www.washingtonpost.com/news/the-switch/wp/2018/06/01/google-to-drop-pentagon-ai-contract-after-employees-called-it-the-business-of-war/; Andrew Eversden, “Army Awards Palantir $823M Contract For Enterprise ‘Data

国家机构难以垄断高精尖军事技术，军火商、互联网科技巨头、人工智能技术公司以及大数据分析机构等多个主体均能够深度参与研发进程。然而，研究与实践表明，减少参与者数量更有利于军控协议的成功。[①] 多元化主体的加入可能增加形成共识的难度，并使得协议面临核查机制和执行程序的复杂化问题，提高违约行为发生的可能性，从而增加军备控制进程的不确定性。

其次，逐利性商业模式加剧技术扩散风险。人工智能技术在商业领域具有巨大的潜力，企业在技术发展中占据优势地位。例如，英伟达公司在图形处理器（GPU）研发领域深耕多年，成为算力行业的龙头企业。再如，OpenAI公司在大语言模型的研发领域展现出惊人的创造力。此类领先优势是政府及其主导机构难以企及的。为了在激烈的市场竞争中保持领先，企业需要快速高效且持续不断地推动技术进步，并实现盈利。在商业利益的驱动下，掌握核心技术的企业往往倾向于投入规模化量产，以迅速降低成本并扩展利润空间。而成本的迅速下降则会造成较高的技术转移与扩散风险，从而不利于军备控制的实施。[②]

由于企业具有的先天技术优势，国家力量无法完全主导人工智能军备的研发进程和治理环节。在多元行为体的深度参与以及商业运作模式的广泛影响下，人工智能技术的持续扩散似乎无法避免，不利于军备竞赛稳定性的建立。

第三，增加美方的军备不透明。一国的军备决策更有可能是威胁感的直接产物，而不一定是威胁的产物。因此，过度的威胁感可能导致对方在军备方面的过度反应，从而增加军备竞赛发生的可能性。[③] 为了消除被夸大的威

Fabric,' " Breaking Defense, October 6, 2021, https://breakingdefense.com/2021/10/army-awards-palantir-823m-contract-for-enterprise-data-fabric/; "Lockheed Martin, Microsoft Announce Landmark Agreement on Classified Cloud, Advanced Technologies for Department of Defense," Microsoft News Center, November 16, 2022, https://news.microsoft.com/2022/11/16/lockheed-martin-microsoft-announce-landmark-agreement-on-classified-cloud-advanced-technologies-for-department-of-defense/.

① [美]肯尼思·华尔兹：《国际政治理论》，信强译，上海人民出版社 2017 年版，第 187 页； Megan Lamberth and Paul Scharre, "Arms Control for Artificial Intelligence," *Texas National Security Review*, Vol. 6, No. 2, 2023, p. 99。

② Christopher F. Chyba, "New Technologies and Strategic Stability," *Dædalus,* Vol. 149, No.2, 2020, pp. 153-154.

③ 李彬：《军备控制理论与分析》，第 197 页。

胁感，防止军备竞赛，当事国需要采取切实措施以提升军备的透明度。既往国际经验表明，两种方法不可或缺。一是采取措施方便对方了解本国的意图，从而避免产生误判。理论研究表明，进攻性意图容易招致对方的强烈反应，而防御性意图更有利于维持稳定局面。[①] 因此，维护军备竞赛稳定性的关键在于向对方表明自身的防御性目的。二是采取硬件方面的信息分享或可核查的手段，使他国明确本国的实力和意图。然而，人工智能军备的特质使得上述方法在执行中面临较大困难。

首先，人类难以判断人工智能的意图。一是由于人工智能的内部运算过程类似于“黑箱”，人类尚难以掌握和解释。现代人工智能模型尤其是深度学习模型通过复杂的数学运算和多层次的抽象来进行学习和决策，使得人类难以直观地理解其内部工作原理。人工智能决策的依据和逻辑不易被人类理解，使得评估其进攻或防御性质较为困难。二是部分人工智能技术本身兼具进攻与防御双重性质。例如，无人机既可用于防御性的情报、监视和侦察，也可用于进攻性的袭击敌方设施。人工智能加持的反导系统通常被认为是防御性的，然而，这些系统可能具备主动探测和攻击功能，甚至服务于“基于预警发射”（launch on warning）的决策模式，[②] 使防御与进攻的界限变得模糊。

其次，对人工智能军备实施有效核查的难度较高。一是人工智能技术具有一定的隐蔽性。人类既往的核查经验多来自对核武器的核查，依赖对核材料、核设施、核试验场以及导弹发射装置等硬件设施进行核查和信息分享。然而，人工智能军备多依赖软件和算法，可以隐藏于常规的算力基础设施中，难以通过传统的军备核查手段发现。二是技术更新迭代迅速。由于商业化运营模式的广泛使用，人工智能技术的更新和改进往往在短期内发生。这种快速变化进一步加大了核查人员准确评估军备能力和用途的难度。三是数据和模型的复杂性。人工智能系统的性能和行为高度依赖于其训练数据和模型参

① Robert Jervis, “Cooperation under the Security Dilemma,” *World Politics*, Vol. 30, No. 2, 1978, pp. 167-214.

② “基于预警发射”指当早期预警系统探测到敌方的导弹来袭迹象时，即在敌方核导弹击中本国之前迅速反击，确保本国的核力量得以运用，以避免敌方先发制人摧毁己方核武库。此种模式虽然具有威慑力，但可能因技术故障或误报而导致误射，具有较高的战略风险。

数。即使核查人员能够访问这些系统，也未必能够理解其复杂的训练过程和参数配置。四是人工智能技术具有一定的欺骗性。为了让人工智能系统能够在复杂现实环境中保持稳定和有效运行的能力，不少系统具有一定的“鲁棒性”（robustness）设计，[①] 导致其在对抗性环境中可能有意隐藏真实意图和能力，从而增加核查人员识别和评估的难度。

由此可见，人类难以仅通过硬件层面的核查或信息分享来掌握人工智能军备的具体情况。然而，军备不透明度可能造成双方的安全疑虑，从而加剧军备竞赛的不稳定性（见图 4）。

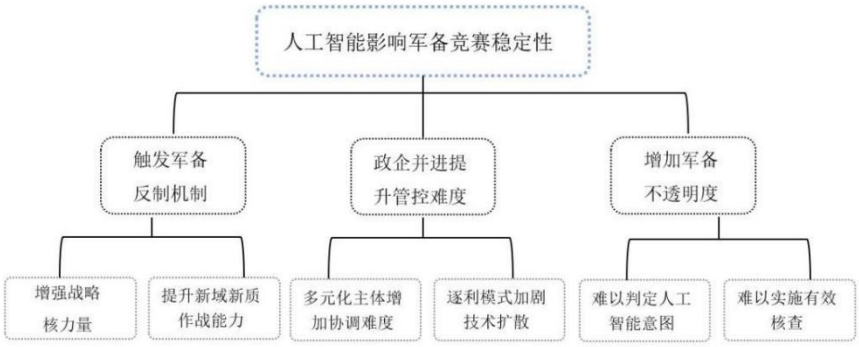

图 4　美国人工智能应用可能冲击军备竞赛稳定性的作用机制

资料来源：作者自制。

四、中国应对美国军备人工智能化的路径

人工智能技术的迅速发展及其在核武器等军事领域的广泛应用前景，使中美大国博弈面临全新形势，并为中美战略稳定关系注入新的不确定因素。由于人工智能军备的特质，既有的军备控制经验难以完全适用于维护战略稳

① “鲁棒性”（robustness）指人工智能系统在面对多样化场景或对抗压力时的抵抗力，特别是保证其目标的正确性以及能力泛化性。参见吉嘉铭等：《人工智能对齐：全面性综述》，北京大学人工智能研究院 AI 安全与治理中心译，2024 年 1 月，第 6 页，https://alignmentsurvey.com/uploads/AI-Alignment-A-Comprehensive-Survey-CN.pdf。

定的新需求。为了切实维护中国的国家安全，营造相对稳定的周边环境，必须从多个层面采取有效措施来积极应对新形势。

（一）提升智能化作战能力

随着军事领域的变革以及智能化步伐的加快，“智权”可能成为继“陆权”“海权”之后影响和决定作战全局的新控制权。有鉴于此，中国须在进一步创新军事理论的基础上，提升应对战略领域智能化作战的理论水平和技术能力，以对美国智能化作战能力实施有效反制。

第一，进一步提升军队智能化作战的理论水平。智能化作战是以人工智能为核心的前沿科技在作战指挥、装备、战术等领域渗透、拓展的必然发展方向。[①] 智能化作战旨在实现作战效能的全面提升和战场优势的最大化，是军事领域发展的潮流和趋势。随着技术的不断进步，智能化作战理论需同步发展和完善，为制定军事战略和战术提供更强大的支持。中国须以这一理论为抓手，把握智能化战争的演进脉搏，洞悉其内在本质，提升全军的理论水平，才能抓住变革的新契机，并使作战能力实现“弯道超车”。

第二，利用智能化技术提升中国军备性能。其一，加强人工智能相关的软硬件与基础设施建设。中国须进一步提高芯片性能、网络通信质量和大规模数据存储能力，以达到提升算力的效果。同时，中国还须开发和优化人工智能算法，提高其训练数据和生成推理结果的能力，从而进一步提升人工智能军备的性能。其二，提升预警系统与指挥控制系统的能力。这有利于延长决策窗口期，同时提高信息分析的准确性，降低因仓促决策、信息错误或技术失误等导致的误判风险，从而增强危机稳定性。其三，提高人工智能与新域新质作战能力的兼容性。中国应运用人工智能技术促进多源信息的处理和分发，提升作战体系的态势感知能力、辅助决策能力以及无人化作战平台的协同作战能力，从而增强中国军事力量的非对称优势、快速反应与灵活部署能力，弥补中国对美军备劣势。

（二）推动建立人工智能军备的全球治理框架

以美国军备人工智能化升级为代表的全球军备人工智能化趋势，已然引

① 沈寿林、张国宁：《认识智能化作战》，《解放军报》2018 年 3 月 1 日，第 7 版。

发了国际社会对人工智能风险问题的高度关注与广泛讨论。目前，该领域还缺乏被广泛接受的全球治理框架，包括中国在内的国际社会成员正致力于通过协商有效管控人工智能引发的安全、法律、伦理、人道主义等风险。[①] 有鉴于此，中国应在国际机制中发挥引领作用，努力提升在该领域的话语权和主导权，从而制约美国利用人工智能谋求绝对军事优势和霸权的行为。

第一，推动制定人工智能军备的限制性原则，切实增强危机稳定性。鉴于人工智能特质以及人机互动过程中出现的问题，中国须尽快推动国际社会制定相关规则，防范意外事件导致局势升级。其一，限制核武领域的完全自主系统。中国可倡导确保人工智能在核武器系统中的应用始终受到严格的人类监督和控制，确保所有关键决策必须由人类操作员进行验证和批准。[②] 其二，制定人工智能军备系统操作指导。既往经验表明，维持系统的高可靠性通常需要成员训练有素、警觉微小事故、信息沟通流畅、通过持续学习和模拟演练增强应对突发事件能力等。[③] 人工智能军备系统的运作应在借鉴既有经验的基础上创新，尽可能保证其安全性与稳定性。其三，推动国际社会各成员国对人工智能军备“实施分级、分类管理，避免使用可能产生严重消极后果的不成熟技术”。[④] 此外，中国还可考虑在国际层面推动有核国家放弃“基于预警发射”模式，降低因误报或技术故障导致的意外发射风险。[⑤]

第二，推动增进人工智能军备透明度和防扩散，提高军备竞赛稳定性。其一，推动国际社会对于人工智能的可解释性和可控性研究。[⑥] 人工智能的

---

① 《中国就“致命性自主武器系统”问题向联合国秘书长提交的文件》，外交部网站，2024 年 5 月 23 日，https://www.mfa.gov.cn/wjb_673085/zzjg_673183/jks_674633/fywj_674643/202405/t20240523_11310587.shtml。

② 联合国框架下的讨论已涉及禁止在无人控制或无人监督的情况下运行致命自主武器系统的倡议。参见 *Our Common Agenda Policy Brief 9: A New Agenda for Peace*, United Nations, July 2023, p. 27, https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-new-agenda-for-peace-en.pdf。

③ Karl E. Weick and Kathleen M. Sutcliffe, *Managing the Unexpected: Sustained Performance in a Complex World*, 3rd edition, San Francisco: Jossey-Bass Inc. Publishers, 2015, pp. 1-128.

④《中国关于规范人工智能军事应用的立场文件》，外交部网站，2021 年 12 月 14 日，https://www.mfa.gov.cn/web/ziliao_674904/zcwj_674915/202112/t20211214_10469511.shtml。

⑤ 关于该倡议的研究参见 Pavel Podvig, “Reducing the Risk of an Accidental Launch,” *Science & Global Security*, Vol. 14, No. 2-3, 2006, pp. 75-115。

⑥ “可解释性”要求人类能够理解人工智能系统的内在推理过程，或构建可解释性工具，

“黑箱”属性是各国担忧人工智能军备可能引发严重后果的重要原因之一。推动上述领域的持续深入研究可在相当程度上减少各国之间因人工智能技术特质而导致的相互疑惧。[①] 其二，推动国际社会吸纳企业机构参与制定人工智能研发的相关监管和治理框架，[②] 加快设立明确的技术管控标准和审查评估机制，努力遏制敏感技术的无序扩散。

（三）尝试拓展中美合作空间

作为国际社会中具有显著影响力的大国，中国和美国在避免人工智能军备引发灾难性后果方面具有一定的共同利益。2021 年 12 月，中国在联合国《特定常规武器公约》第六次审议大会上呼吁，各国应通过对话与合作，就如何规范人工智能军事应用寻求共识，构建有效治理机制，避免人工智能军事应用给人类带来重大损害甚至灾难。[③] 美国亦认为，有必要与中、俄等“竞争对手”讨论人工智能对危机事件的影响，并防止冲突升级。[④] 目前，中美两国已在该问题上表现出一定程度的合作意向。[⑤] 双方应在共同利益的基础上，采取切实措施保证人工智能军备的安全性与稳定性，并围绕危机防范与军备竞赛管控等议题探索对话与合作的可能性。

第一，中美可在防范危机升级方面采取多重措施。其一，共同承诺维持由人类控制核武器的使用，[⑥] 考虑在所有核武器发射决策中引入多层次人类

深入了解神经网络内部的概念和推理机制。“可控性”则要求确保系统的行动和决策过程始终受到人类监督和约束，以保证人类可以及时纠正系统行为中的任何偏差或错误。参见吉嘉铭等：《人工智能对齐：全面性综述》，第 6 页。

① 我国已经提出相关倡议，参见《发展负责任的人工智能：新一代人工智能治理原则发布》，科技部网站，2019 年 6 月 17 日，https://www.most.gov.cn/kjbgz/201906/t20190617_147107.html。

② 《抓住安全、可靠和值得信赖的人工智能系统带来的机遇，促进可持续发展》，联合国网站，2024 年 3 月 11 日，第 3 页，https://documents.un.org/doc/undoc/ltd/n24/065/91/pdf/n2406591.pdf。

③ 《中国首次就规范人工智能军事应用问题提出倡议》，新华网，2021 年 12 月 14 日，http://www.news.cn/world/2021-12/14/c_1128160251.htm。

④ *Final Report*, U.S. National Security Commission on Artificial Intelligence, 2021, p. 10, https://irp.fas.org/offdocs/ai-commission.pdf.

⑤ 《习近平同美国总统拜登举行中美元首会晤》，外交部网站，2023 年 11 月 16 日，https://www.mfa.gov.cn/web/zyxw/202311/t20231116_11181125.shtml。

⑥ 该倡议在 2024 年中美元首会晤中已有提及。参见《外交部发言人全面介绍中美元首利马会晤情况》，外交部网站，2024 年 11 月 17 日，https://www.mfa.gov.cn/fyrbt_673021/202411/t20241117_11527684.shtml。

监督和控制程序，尽量避免自动化和算法决策可能导致的意外状况。其二，共同承诺对相关武器系统实施严格的测试和验证流程，确保其在各种可能的使用场景下表现出预期的行为，以避免误判和误操作。其三，完善危机沟通热线和应急通报机制，推动实施敏感行动的提前通报以缓解对方疑虑，尽量避免因误判等引发的危机升级。

第二，中美可探讨管控军备竞赛的切实举措。其一，可考虑促成双方就不使用不成熟技术、不对核武人工智能系统实施对抗性攻击等达成共识，推动双方在军备控制领域建立信心。其二，两国可考虑采取二轨或“1.5 轨”对话等方式，推动双方建立高级别定期会晤与沟通机制，发挥这些机制在该问题领域交流信息和澄清意图的作用，使军备的发展和竞争更具可控性。

结　　语

在国际权力格局转变以及人工智能技术快速发展的大背景下，美国不遗余力地推动军备领域的人工智能化进程，甚至放任人工智能技术在核武等关键领域不断渗透，以维持其在全球范围内的军事优势和霸权地位。考虑到人工智能对于军备力量的显著赋能及其固有特质，美国军备领域的人工智能化进程不仅将整合美国军事力量的构成，还将对中美战略稳定产生负面冲击。

而从更加宏观的视角来看，军备人工智能化的影响还远不止于此。一是国际格局的重塑。人工智能或将重新定义国际安全框架，改变传统的军事战术和战略，迫使各国重新评估和调整其国防政策和安全策略。而新的技术优势可能导致部分国家在权力竞争中获得优势地位，从而显著改变国际地缘政治格局。二是法律和伦理议题的挑战。例如，人工智能的决策伦理、战争责任归属、审查和监督以及防范技术滥用等问题，都亟须各国的协商和合作。为了应对人工智能化浪潮所带来的复杂局势，国际社会推动形成一个被广泛接受的人工智能军备治理框架势在必行。为实现这一目标，中美等大国应秉持负责任态度进行自我约束，努力减少技术滥用风险，并探索在该领域的合作空间，从而促进全球安全和稳定。

[责任编辑：张　珺]