致命性自主武器系统军控*

——困境、出路和参与策略

龙　坤　　徐能武

【内容摘要】 致命性自主武器系统对国际安全和人道主义构成重大挑战，日益成为国际社会关注的重要议题。当前，致命性自主武器系统军控仍然面临概念不清、政治障碍、军事诱惑及问责困难等诸多问题，难以实质性推进。基于对致命性自主武器系统概念、原则和机制的分析，国际社会宜从现有国际法和军控体系出发，以联合国《特定常规武器公约》机制为主导，厘清主要概念，遵守明确区分原则、比例原则，健全相关机制，推动致命性自主武器系统军控稳步前行。中国需重视参与和推动这一议题的探讨，提升在致命性自主武器系统军控谈判中的话语权；同时，中国也要树立底线思维，做好致命性自主武器系统军控受挫的相关准备，加快推进中国特色军民融合式人工智能发展进程。

【关键词】 致命性自主武器系统　军备控制　《特定常规武器公约》　联合国

【作者简介】 龙坤，国防科技大学文理学院博士研究生，国防科技战略研究智库实习研究员（长沙　邮编：410073）；徐能武，国防科技大学文理学院教授（长沙　邮编：410073）

【中图分类号】 D815.1　　**【文献标识码】** A

【文章编号】 1006-1568-(2020)02-0078-25

【DOI 编号】 10.13851/j.cnki.gjzw.202002005

* 本文系国际军事合作办公室 2018 年度项目“致命性自主武器系统军控研究”的阶段性成果。

随着人工智能技术的不断进步，无人军事平台也在迅速发展，相关技术的武器化带来了国际安全和人道主义隐忧，引起了国际社会的广泛关注。人工智能全球治理逐步提上了议程，[①] 尤其引人注目的是致命性自主武器系统（Lethal Autonomous Weapons Systems, LAWS）及其控制。2013 年以来，致命性自主武器系统相关议题在国际社会引起越来越广泛的重视和争议，针对致命性自主武器系统进行军备控制的呼声也日益高涨。目前，已有不少国内外学者对致命性自主武器系统可能带来的法律、安全、伦理等问题进行了研究，[②] 但尚未有学者深入分析致命性自主武器系统军控当下面临的主要困境，亦未提出相应的中国方案。在此背景下，本文试从概念、原则、制度和监管四个方面提出推动致命性自主武器系统军控的基本框架，认为应从现有国际法和军控体系出发，以联合国《特定常规武器公约》机制为主导，厘清相关概念，明确区分原则、比例原则，逐步编织和完善制度之网，健全监管和核查机制，推动致命性自主武器系统军控稳步前行。

对中国而言，需要重视参与和推动这一议题的探讨，明确自身的核心观点，提升中国在致命性自主武器系统军控谈判中的话语权。同时，也要树立底线思维，做好致命性自主武器系统军控受挫的相关准备，加快推进中国特色军民融合式人工智能革命。

① 陈伟光、袁静：《人工智能全球治理：基于治理主体、结构和机制的分析》，《国际观察》2018 第 4 期，第 23—37 页；巩辰：《全球人工智能治理——“未来”到来与全球治理新议程》，《国际展望》2018 第 5 期，第 36—55 页。

② 代表性文献有，刘杨钺：《全球安全治理视域下的自主武器军备控制》，《国际安全研究》2018 年第 2 期，第 49—71 页；董青岭：《新战争伦理：规范和约束致命性自主武器系统》，《国际观察》2018 年第 4 期，第 51—66 页；徐能武、葛鸿昌：《致命性自主武器系统及其军控选择》，《现代国际关系》2018 年第 7 期，第 54—62 页；徐能武、龙坤：《联合国 CCW 框架下军控辩争的焦点与趋势》，《国际安全研究》2019 年第 5 期，第 108—132 页；Nathan Leys, “Autonomous Weapon Systems and International Crises,” *Strategic Studies Quarterly*, Vol. 12, No. 1, 2018, pp. 48-73; Frank Sauer, “Autonomous Weapon Systems and Strategic Stability,” *Survival*, Vol. 59, No. 5, 2017, pp. 117-142; Patrick Lin, George Bekey, and Keith Abney, *Autonomous Military Robotics: Risk, Ethics, and Design*, U.S. Department of Navy, Office of Naval Research, December 20, 2008, http://ethics.calpoly.edu/ONR_report.pdf; Andrew P. Williams and Paul Scharre eds., *Autonomous Systems: Issues for Defence Policymakers*, The Hague: NATO Communications and Information Agency, 2015; Paul Scharre, *Autonomous Weapons and Operational Risk*, Ethical Autonomy Project, Center for a New American Security, February 2016, https://www.files.ethz.ch/isn/196288/CNAS_Autonomous-weapons-operational-risk.pdf; Paul Scharre, *Army of None: Autonomous Weapons and the Future of War*, New York: W. W. Norton & Company, 2018.

一、人工智能革命背景下的致命性自主武器系统军控

致命性自主武器系统军备控制是在 2013 年才进入国际军备控制领域的一个新兴议题，其直接原因是人工智能的快速发展并不断走向军事化、武器系统的自主性不断提升，对于国际安全和人道主义带来了潜在威胁，引起了国际社会广泛重视。

（一）致命性自主武器系统军控的相关概念

一般而言，自主武器（autonomous weapons）被界定为能独立完成搜索目标、决定打击目标以及实施打击等全部作战任务周期的武器系统。[①] 相比自主武器，自主武器系统则是指包括传感器、决策单元和弹药等整个构成的能自主完成作战任务的武器系统。而致命性自主武器系统则是指具有致命杀伤力的自主武器系统，也被称为“杀手机器人”（killer robots）或“致命性自主机器人”（lethal autonomous robots, LAR）。[②] 从历史上来看，武器系统的自主性呈现一种不断提升的趋势，武器系统进步的同时也伴随着自主程度的提升。据统计，目前至少有 90 个国家拥有无人机，16 个国家或非国家行为体拥有武装无人机。[③] 美国 2018 年 2 月发布的《核态势评估报告》披露，俄罗斯正在开发一种名为“斯塔图斯-6 号”（Status 6）的“新型洲际核动力海底自主鱼雷”；此外，俄罗斯还在积极研发智能导弹、无人机、无人车和军用机器人。[④] 需要指出的是，目前关于致命性自主武器系统的概念和特点，国际上仍然存在很大争议，没有形成共识。

① [美]保罗•沙瑞尔：《无人军队：自主武器与未来战争》，朱启超、王姝、龙坤译，世界知识出版社 2019 年版，第 57—58 页。

② 这几种说法几乎是同一个意思，一般情况下可以通用。

③ Paul Scharre, “Killer Robots and Autonomous Weapons with Paul Scharre,” *Center for A New American Security*, https://www.cnas.org/publications/podcast/killer-robots-and-autonomous-weapons-with-paul-scharre.

④ Boris Egorov, “Rise of the Machines: A Look at Russia’s Latest Combat Robots,” Russia Beyond, June 6, 2017, https://www.rbth.com/defence/2017/06/06/rise-of-the-machines-a-look-at-russias-latest-combat-robots_777480; and Nikolai Litovkin, “Comrade in Arms’: Russia is Developing a Freethinking War Machine,” August 9, 2017, https://www.rbth.com/defence/2017/08/09/comrade-in-arms-russia-is-developing-a-freethinking-war-machine_819686.

军备控制是指对武器及其相关设施、相关活动或者相关人员进行约束，例如对军备的研究、生产、存储、部署、使用方式等进行约束。[①] 冷战期间，学者们对以核武器为代表的战略军备控制进行了广泛而深入的研究，逐渐形成了以战略稳定性为核心概念的经典军备控制理论。冷战结束后，随着科技的迅速发展，军备控制的领域也在不断拓展，较有代表性的有网络空间、外层空间等新兴领域的军备控制。[②] 如今，随着人工智能的迅猛发展和不断走向军事领域，致命性自主武器系统被认为可能会威胁到国际安全与全球战略稳定，因此也迅速进入了国际军控界的视野。研究显示，美国、以色列、韩国、英国、俄罗斯等国正开发在选择和攻击目标的关键功能方面具有相当程度自主性的武器系统，如果不加以控制，世界可能会陷入自主武器军备竞赛的危险境地，破坏世界和平与稳定。[③] 顾名思义，致命性自主武器系统军备控制就是指对致命性自主武器及其相关设施、相关活动或者相关人员进行约束和控制，尤其是指对于致命性自主武器系统的研究、发展、部署和使用进行限制和约束。

（二）推动致命性自主武器系统军控的主要平台

目前，致命性自主武器系统军控已经成为全球安全治理的热点话题，出现了多样化的平台。归纳起来，主要有以下几种。

最核心和正式的官方平台当属联合国《特定常规武器公约》（Convention on Certain Conventional Weapons, CCW）会谈机制。该会谈机制致力于“禁止或限制使用被认为对战斗人员造成过分杀伤或不必要痛苦或波及平民的特定类型武器”；此前曾讨论过对无法检测的碎片、地雷、燃烧武器、激光致盲武器、战争遗留爆炸物等常规武器的军备控制问题，并通过了相应的五项议定书；[④] 从 2014 年至今，针对致命性自主武器系统这一议题，该会谈

① 李彬：《军备控制理论与分析》，国防工业出版社 2006 年版，第 4 页。

② 关于网络空间军备控制的相关研究可参见程群：《网络军备控制的困境与出路》，《现代国际关系》2012 年第 2 期，第 15—21 页；关于太空军备控制的经典文献可参见夏立平：《外层空间军备控制的进展与障碍》，《当代亚太》2002 年第 6 期，第 35—40 页；聂资鲁：《外层空间军备控制与国际法》，《甘肃政法学院学报》，2007 年第 4 期，第 95—103 页。

③ 参见官网：The Campaign to Stop Killer Robots, https://www.stopkillerrobots.org/learn/。

④ 关于该机制的历史沿革以及附加议定书的主要内容，可参见：Convention on Certain Conventional Weapons, Prohibits or Restricts the Use of Conventional Weapons Which Are

机制已经召开了三次非正式专家会议，并成立专门的政府专家组（Group of Governmental Experts on Lethal Autonomous Weapons Systems, GGE on LAWS）围绕该议题召开了多次会议。[①] 最近的一次是 2019 年 3 月和 8 月召开的政府专家组会议，讨论通过了《致命性自主武器系统领域的新技术问题政府专家组 2019 年会议报告》，确立了多项指导原则。[②] 除了这一机制外，联合国裁军事务厅（United Nations Office for Disarmament Affairs）、联合国裁军研究所（UN Institute for Disarmament Research）等联合国附属机构也在积极推动这一议题的探讨和解决。[③]

与此同时，红十字国际委员会（International Committee of the Red Cross, ICRC）也汇集了多个领域的专家讨论致命性自主武器系统可能带来的技术、法律、伦理等问题，尤其聚焦如何用现有国际人道法规制自主武器系统，并形成报告递交《特定常规武器公约》会议讨论。[④] 除此之外，联合国框架下的人权理事会与联合国大会第一委员会（裁军与国际安全委员会）也是重要的探讨平台。自 2013 年以来，致命性自主武器系统就成为联合国大会第一委员会每年都要讨论的重点议题，越来越多的国家对该问题表示关切。

除了联合国等官方平台之外，也出现了很多致力于推动致命性自主武器

Considered Excessively Injurious or Whose Effects Are Indiscriminate, https://unoda-web.s3-accelerate.amazonaws.com/wp-content/uploads/assets/publications/more/ccw/ccw-booklet.pdf。

① 有关该会谈机制下致命性自主武器系统军控探讨焦点、分歧和未来趋势的总结和初步研究，可参见：徐能武、龙坤：《联合国 CCW 框架下致命性自主武器系统军控辩争的焦点与趋势》，《国际安全研究》2019 年第 5 期，第 108—132 页。也可参见 UN CCW on LAWS 官网：https://www.unog.ch/80256EE600585943/(httpPages)/8FA3C2562A60FF81C1257CE600393DF6?OpenDocument。

② CCW, *Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, September 25, 2019, https://www.unog.ch/80256EE600585943/(httpPages)/5535B644C2AE8F28C1258433002BBF14?OpenDocument.

③ 参见联合国这几个下属机构的官网：https://www.un.org/disarmament/publications/occasionalpapers/unoda-occasional-papers-no-30-november-2017/; https://www.unidir.org/files/medias/pdfs/unidir-statement-ccw-expert-meeting-laws-april-2018-eng-0-746.pdf。

④ "Autonomous Weapons Systems: Technical, Military, Legal, and Humanitarian Aspects," Expert Meeting, International Committee of the Red Cross, March 28, 2014, https://www.icrc.org/en/download/file/1707/4221-002-autonomous-weapons-systems-full-report.pdf, and "Autonomous Weapons Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons," Expert Meeting, International Committee of the Red Cross, March 16, 2016, https://www.icrc.org/en/download/file/21606/ccw-autonomous-weapons-icrc-april-2016.pdf.

系统军控的非政府组织。2009 年，诺埃尔·夏基（Noel Sharkey）等学者共同成立了“国际机器人军备控制委员会”（International Committee for Robot Arms Control, ICRAC），旨在呼吁国际社会对致命性自主武器系统进行广泛探讨和军备控制，尤其是要禁止研发和使用具备核杀伤力的自主武器以及太空自主武器。[①] 2012 年 10 月，“禁止杀手机器人运动”（Campaign to Stop Killer Robots）成立，作为一个非政府组织全球联盟，它致力于禁止使用完全自主武器，从而保持人类对使用武力的实际控制。截至 2020 年 1 月，已有来自 61 个国家共 140 个国际、区域和非政府组织加入该运动。由于其众多的参与方和广泛的影响力，该运动也成为国际上推动致命性自主武器军备控制最为活跃的非官方力量。[②] 此外，人权观察组织、未来生活研究所等非政府组织也积极参与了推动限制和禁止致命性自主武器系统的相关运动。[③]

（三）参与致命性自主武器系统军控的主要行为体

目前，围绕致命性自主武器系统军控，国际上也出现了多元参与主体。

一是主权国家。如前所述，越来越多的主权国家意识到了致命性自主武器系统的重要性，积极参与到诸如联合国《特定常规武器公约》、人权理事会、“禁止杀手机器人运动”等会议和运动中来，表明本国对于致命性自主武器系统的立场和态度。最为突出的就是《特定常规武器公约》会谈机制已围绕该议题召开多次正式会议并成立了专门的政府专家组，吸引了近百个国家的参与。[④]

二是联合国等国际组织人员。较有代表性的有联合国人权理事会“法外处决、即审即决或任意处决问题”特别报告员克里斯托弗·海恩斯（Christof Heyns），他在 2013 年就向联合国大会提交报告，呼吁重视分析和解决以武

① 参见国际机器人军备控制委员会官网，https://www.icrac.net/about-icrac/。

② 参见“禁止杀手机器人运动”官网，https://www.stopkillerrobots.org/members/。

③ 参见人权观察组织对于 LAWS 的表态：Human Rights Watch Statement to the CCW Group of Governmental Experts Options for Future Work on LAWS, March 27, 2019，https://www.unog.ch/80256EDD006B8954/(httpAssets)/41D084E97FF20C84C12583CB003EE0B5/$file/CCW+HRW+statement+27+March+19.pdf；及未来生活研究所官网：https://futureoflife.org/。

④ 详见徐能武、龙坤：《联合国 CCW 框架下致命性自主武器系统军控辩争的焦点与趋势》，第 108—132 页。

装无人机为代表的致命性自主武器系统对国际人道法和人权造成的冲击问题，很多国家也对这一报告表示了关切。[①]

三是知名学者和科学家群体。比较有代表性的有史蒂芬·霍金（Stephen Hawking）、诺埃尔·夏基等。例如，霍金就警告人工智能的快速发展可能导致人类灭亡。[②] 2015 年 7 月，1 000 多名人工智能专家联名发表了一封公开信，严重警告人工智能军备竞赛的危险，并呼吁禁止使用自主武器系统。信中指出："即使自主武器的部署不合法，但在几年内就可能成为现实，自主武器将在火药和核武器之后引发第三次战争革命。今天人类面临的关键问题是启动全球人工智能军备竞赛还是将它阻止在萌芽状态。"[③] 2017 年，100 余名机器人和人工智能公司领导人再次联名发表了一封公开信，警告人们自主武器系统造来的危险。[④]

四是商界领袖和智库人员。商界代表人物有特斯拉创始人埃隆·马斯克（Elon Musk）、苹果联合创始人史蒂夫·沃兹尼亚克（Steve Wozniak）等。马斯克将发展通用人工智能比作"召唤恶魔"，警告各国竞相发展人工智能武器可能导致全球军备竞赛，呼吁禁止攻击性自主武器。[⑤] 智库也是重要的参与方，比较有代表性的有美国智库新美国安全中心（Center for a New American Security）、斯德哥尔摩国际和平研究所等。[⑥] 其中，美国智库新美

① Christof Heyns, "Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions," September 13, 2013, http://www.justsecurity.org/wp-content/uploads/2013/10/UN-Special-Rapporteur-Extrajudicial-Christof-Heyns-Report-Drones.pdf?utm_source=Press+mailing+list&utm_campaign=6de0426c90-2013_10_17_Heyns_drones_report_UN&utm_medium=email&utm_term=0_022da08134-6de0426c90-286021377。各国具体关切可参见 Report on Outreach on the UN Report on Lethal Autonomous Robotics, July 31, 2013, http://www.stopkillerrobots.org/wp-content/uploads/2013/03/KRC_ReportHeynsUN_Jul2013.pdf。

② Tanya Lewis, "Stephen Hawking: Artificial Intelligence Could End Human Race," *Live Science*, December 2, 2014, https://www.livescience.com/48972-stephen-hawking-artificial-intelligence-threat.html.

③ "Autonomous Weapons: An Open Letter From AI & Robotics Researchers," *Future of Life Institute*, July 2015, http://futureoflife.org/open-letter-autonomous-weapons/.

④ "An Open Letter to the United Nations Convention on Certain Conventional Weapons,"*Future of Life Institute*, 2017, https://futureoflife.org/autonomous-weapons-open-letter-2017/?cn-reloaded=1.

⑤ Eric Mack, "Elon Musk: "We Are Summoning the Demon' With Artificial Intelligence," CNET, October 26, 2014, https://www.cnet.com/news/elon-musk-we-are-summoning-the-demon-with-artificial-intelligence/.

⑥ 代表文献有：Vincent Boulanin and Maaike Verbruggen, *Mapping The Development Of Autonomy In Weapon Systems*, Stockholm International Peace Research Institute, November 2017,

国安全中心的保罗·沙瑞尔（Paul Scharre）长期关注这一议题，并发表了关于致命性自主武器的一系列作品和言论。[①] 综合来看，这些参与致命性自主武器军控的群体采取的主要方法有公开呼吁、发表公开信、召开会议等。与以往军备控制不同的是，这次针对致命性自主武器系统军控的探讨吸引了主权国家、国际组织、实体机构、商界领袖、学界和智库精英等多个群体和利益攸关方的广泛关注和积极参与，其重要性和复杂性可见一斑。

二、致命性自主武器系统军控面临的主要困境

当前，推进致命性自主武器系统军控尤其是禁止杀手机器人已经形成了一股国际潮流。根据“禁止杀手机器人运动”网站的统计，截至 2019 年 8 月，已经有 29 个国家明确表示支持禁止完全自主武器系统。[②] 尤为引人注目的是，联合国《特定常规武器公约》会谈机制在推动致命性自主武器系统军控上发挥了重要作用，形成了一系列报告文件。

但是，迄今为止国际上并没有对致命性自主武器系统的定义形成共识，也没有形成具有法律约束力的文件来规范该武器系统的发展控制和使用。换言之，致命性自主武器系统军控并没有取得实质性进展，外交步伐没有赶上其技术的发展速度。[③]

（一）致命性自主武器系统的相关概念存在广泛争议

对于致命性自主武器系统的相关概念，各国目前尚未形成足够共识。比如，什么才能称得上是致命性自主武器？是彻底禁止它的研发、生产和使用，

---

https://sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf。

① 代表作品有：Paul Scharre, *Army of None: Autonomous Weapons and the Future of War*, New York: W. W. Norton & Company, 2018; Paul Scharre, *Autonomous Weapons and Operational Risk* 等。

② “Country Views on Killer Robots,” Campaign to Stop Killer Robots, https://www.stopkillerrobots. org/wp-content/uploads/2019/08/KRC_CountryViews21Aug2019.pdf. 这 28 个国家包括巴基斯坦、厄瓜多尔、埃及、古巴、加纳、玻利维亚、津巴布韦、阿尔及利亚、哥斯达黎加、墨西哥、智利、巴拿马、秘鲁、阿根廷、委内瑞拉、巴西、伊拉克、乌干达、奥地利、中国、吉布提、哥伦比亚、摩洛哥、约旦等。

③ 关于目前致命性自主武器军控中的主要分歧可参见徐能武、龙坤：《联合国 CCW 框架下军控辩争的焦点与趋势》，第 108—132 页。

还是只限制它的某一个环节？致命性自主武器对于平民来说究竟是福音还是祸患？以“自主”这一概念为例，“自主性”（autonomy）在不同的领域有着不同的含义。工程学意义上的自主性通常是指机器能在无人干预的情况下独立运行；哲学意义上的自主性主要指道德独立；政治学上自主性的内涵更接近于自治，即自我管理。在军事装备领域，自主性的内涵存在很大争议，究竟什么样的武器系统才能称得上是“自主的”？不同学者、不同国家界定不一。一些机构对于自主武器的界定比较严苛。例如，英国国防部将自主武器系统定义为“能够理解高级意图和方向的系统，基于这种理解及其对环境的感知，这种系统能够采取适当的行动来实现自身目标。”① 而一些学者对自主武器的界定标准则相对比较低，例如，彼得·阿萨罗（Peter Asaro）和马克·古布鲁德（Mark Gubrud）就认为，任何能够在没有人类的操作、决定或确认的情况下释放致命武力的武器系统都可被视为是自主的。古布鲁德更是进一步认为，武器系统不需要能够完全自己做出决定，只要它主动涉及从寻找目标到最终攻击这一“准备过程”的一个或多个部分，它就可以被认为是自主的。② 正是由于这种定义分歧的存在，推进致命性自主武器军控面临重重困难。

（二）各国在推进致命性自主武器系统军控问题上的政治意愿不同

历史上，针对特定武器进行军备控制的成功案例（如核武器和生化武器）往往由大国进行主导和推进，各方尤其是大国的政治意愿对于军备控制的成败起着重要作用。但在致命性自主武器系统军控这一问题上，各国由于在人工智能技术发展状况和国家利益的不同而存在明显分歧。总体而言，目前强烈要求限制致命性自主武器系统的主要是中小国家和发展中国家，大国对于致命性自主武器系统军控的意愿明显不足，认为目前致命性自主武器系统的利害尚不明晰，草率制定禁止政策有些为时过早。例如，2018 年 4 月，参

① U.K. Ministry of Defense, “Joint Doctrine Publication 0-30. 2 Unmanned Aircraft Systems,” https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/673940/doctrine_uk_uas_jdp_0_30_2.pdf.

② Peter Asaro, “On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making,”*International Review of the Red Cross*, June 2012, https://international-review.icrc.org/articles/banning-autonomous-weapon-systems-human-rights-automation-and-dehumanization-lethal.

加联合国《特定常规武器公约》会议的绝大多数国家代表建议制定一项具有法律约束力的关于完全自主武器的文书即议定书或条约，但法国、以色列、俄罗斯、英国和美国五个国家都表示反对。[①] 在同年 8 月的会议上，与会各方再次讨论了是否需要就完全自主武器问题展开正式谈判，并倡议制定一个全面禁止完全自主武器的条约。但是，由于美国、俄罗斯、韩国、以色列和澳大利亚等国的反对，这个提案最终未能通过。[②] 截至 2019 年 8 月，有 12 个国家反对制定一项禁止致命性自主武器系统的国际条约。[③] 以美国为代表的发达国家认为，非但不应禁止致命性自主武器系统研发，反而要鼓励相关的技术创新，原因在于它有望带来很多军事利好和人道主义红利，比如能够增强对平民及民用目标的军事感知能力，在战争中降低平民伤亡的风险。[④] 而其他国家则针锋相对地认为，致命性自主武器系统无法区分军人和平民，可能带来问责真空，降低战争门槛，严重威胁平民安全和人的尊严，不应放任其发展。因此，推进致命性自主武器系统军控也存在明显的政治障碍。

（三）人工智能的潜在军事价值对各国诱惑巨大

目前，世界各国尤其是军事大国正在或明或暗推进人工智能的军事化，这很大程度上是因为人工智能在情报分析、指挥控制、网络作战、军事后勤等领域具有很大的军事应用前景，[⑤] 能够显著提升传统武器系统的自主性，而高自主性意味着利用机器能够实现远超人类的计算及反应速度，并在人机

① “Convergence on Retaining Human Control of Weapons Systems,” April 13, 2018, https://www.stopkillerrobots. org/2018/04/convergence/. 详细报告参见: Report on Activities：Convention on Conventional Weapons Group of Governmental Experts meeting on Lethal Autonomous Weapons Systems, April 2018, https://www.stopkillerrobots.org/wp-content/uploads/2018/07/KRC_ReportCCWX_Apr2018_UPLOADED.pdf.

② 《美俄坚决反对！联合国全面禁止 AI 自主武器正式谈判无果而终！》，新浪网，2018 年 9 月 2 日，https://tech.sina.com.cn/roll/2018-09-02/doc-ihinpmnr7963431.shtml。

③ 这 12 个国家分别是澳大利亚、比利时、法国、德国、以色列、韩国、俄罗斯、西班牙、瑞典、土耳其、美国和英国。参见：“Country Views on Killer Robots,” Campaign to Stop Killer Robots, https://www.stopkillerrobots.org/wp-content/uploads/2019/08/KRC_CountryViews21Aug2019.pdf。

④ CCW, *Report of the 2015 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)*, p.16; CCW, “Humanitarian Benefits of Emerging Technologies in the Area of Lethal Autonomous Weapon System,” March 28, 2018, https://www.unog.ch/80256EDD006B8954/(httpAssets)/7C177AE5BC10B588C125825F004B06BE/$file/CCW_GGE.1_2018_WP.4.pdf.

⑤ Kelley M. Sayler, “Artificial Intelligence and National Security,” Congressional Research Service, January 30, 2019, https://crsreports.congress.gov/product/pdf/R/R45178.

通讯受阻时仍然能保持战场信息优势。换言之，战争压力可能会促使各国军队研发和使用致命性自主武器系统。对于大国来说，这是抢占未来军事制高点的战略技术。对于小国来说，这也是有望改变战争游戏规则的战略技术和武器。对于恐怖分子来说，这更是实施恐怖活动的“天赐之物”。迄今为止，美国国防部已经发布了《2018 国防部人工智能战略概要》等一系列人工智能战略文件，[①] 成立“算法战跨职能小组”“联合人工智能中心”和人工智能国家安全委员会等机构，全力推进人工智能的军事应用。[②] 虽然美国国防部原副部长罗伯特·沃克（Robert Work）曾表示美军“不会将做出致命决策的权力委托给机器”，但他也指出可能会重新考虑这一问题，因为“专制国家”和恐怖分子可能会这样做，美国已经面临人工智能领域的“斯普特尼克时刻”。[③] 与此同时，俄罗斯总统普京强调主导人工智能的国家将成为未来世界的领导者，[④] 俄罗斯也正在着手组建专门开发军用人工智能应用的机构。2017 年，俄罗斯成立了高级研究基金会，专门研究自主能力和机器人技术，为此还召开了一场以“俄军自动化”为议题的年度会议。[⑤] 2018 年 3 月，俄罗斯政府发布了人工智能 10 项议程，计划组建国家人工智能中心等机构。[⑥] 从目前的情况来看，驱动大国制造致命性自主武器系统的军事诱惑

① U.S. Department of Defense, *Summary Of The 2018 Department Of Defense Artificial Intelligence Strategy*, February 11, 2019, https://media.defense.gov/2019/Feb/12/ 2002088963/ -1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF, p.4.

② Sydney J. Freedberg, Jr., “Pentagon Rolls Out Major Cyber, AI Strategies This Summer,” *Breaking Defense*, July 17, 2018, https://breakingdefense.com/2018/07/pentagon-rolls-out-major-cyber-ai-strategies-this-summer/; and P. L. 115232, Section 2, Division A, Title X, §1051.

③ Dan Lamothe，“Pentagon Examining the ‘Killer Robot’ Threat,” *Boston Globe*, March 30, 2016, https://www.bostonglobe.com/news/nation/2016/03/30/the-killer-robot-threat-pentagon-examining-how-enemies-could-empower-machines/sFri6ZDifwIcQR2UgyXlQI/story.html; Colin Clark, “Our Artificial Intelligence ‘Sputnik Moment’ is Now: Eric Schmidt and Bob Work,” *Breaking Defense*, November 1, 2017, https://breakingdefense.com/2017/ll/our-artificial-intelligence-sputnik-moment-is-now-eric-schmidt-bob-work/; and Robert Work, Remarks at the Atlantic Council Global Strategy Forum, Washington, DC, May 2, 2016, https://atlanticcouncil.org/event/2016-global-strategy-forum/.

④ James Vincent, “Putin Says the Nation that Leads in AI Will be the Ruler of the World,” *The Verge*, September 4, 2017, https://www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world.

⑤ Samuel Bendett, “Red Robots Rising: Behind the Rapid Development of Russian Unmanned Military Systems,” *The Strategy Bridge*, December 12, 2017, https://thestrategybridge.org/the-bridge/2017/12/12/red-robots-rising-behind-the-rapid-development-of-russian-unmanned-military-systems.

⑥ Samuel Bendett, “Here’s How the Russian Military Is Organizing to Develop AI,” *Defense*

和动力远大于对其实行军控的考虑。

（四）技术和问责问题

除了以上三大问题之外，在技术和问责方面还存在着诸多难题，首先是区分难题。由于致命性自主武器系统概念界定存在争议，加之自主程度的边界很难划分，这给军控带来了一个很大的技术难题。如果无法明确区分全自主、半自主和自动致命性武器，而一概禁止所有这些在致命性武器系统设计中融入了自动化成分的武器系统，这就像要求消灭战争一样不现实。即使实施了禁令，也难以产生实际效果。其次是问责困境。致命性自主武器系统还可能带来“问责空白”（Accountability Gap）的困境，即可能会出现没有哪一方能对自主武器造成的事故后果负责的窘境。比如，军人如果在作战中使用了自主选择杀戮目标的自主武器误伤了平民，犯了违反战争法或国际人道法的罪行，那么这个责任究竟应该归咎于指挥官，还是操作人员，亦或是程序设计者？倘若这并非是操作员有意为之，而是机器没能按照操作员的意图行事造成的非预期事故（unintended accidents），则会使得问责变得更加困难。此外，致命性自主武器系统还可能会让各方都有推卸责任的理由，因为其背后有多种人员群体，每个人只是造成“事故”的一小部分。因此，人们的罪恶感可能被稀释，每个人都有理由规避自己的责任。[①] 再次是核查困难。核查是军控的难点和关键所在，如果不能有效核查，会严重侵蚀军控条约的可信度和有效性。致命性自主武器系统核查最大的困难在于划清人工智能技术的军民界限。从根本上来讲，人工智能算法是软件而不是硬件，它具有很强的易复制性和扩散性。理论上，人工智能技术采用的机器学习算法和大数据都可以几乎为零的成本进行复制和扩散，应用于不同的场景。比如自动驾驶技术既能用于民用交通工具，也能用于军用无人机、无人车和无人潜航器，人脸识别算法既能用于寻找失踪儿童，也能成为致命性自主武器系统的“火眼金睛”。换言之，人工智能技术具有很强的军民两用特性，军用的致命性

One, July 20, 2018, https://www.defenseone.com/ideas/2018/07/russian-militarys-ai-development-roadmap/149900/.

① 关于自主武器问责困境的详细论述，可参见[美]保罗•沙瑞尔：《无人军队：自主武器与未来战争》，第 294—296 页。

自主武器系统与民用的人工智能设备的支撑技术本质上都是一样的，民用人工智能技术能够通过一定手段转化为军事用途，因此要建立对于致命性自主武器的有效核查机制面临技术上的很大困难。如果被核查国在接受核查时将技术藏匿于民用产品中，待核查之后很快就能转为军用致命性自主武器系统，那么核查就会形同虚设，这也是致命性自主武器系统军控的一大技术难点，并成为很多国家反对在该领域实行军控的重要理由。

三、推动致命性自主武器系统军控的基本路径

鉴于目前致命性自主武器系统军控存在上述困境，难以实质性推进。从理论和实践两个层面出发，笔者试从概念、原则和机制等方面提出推动致命性自主武器系统军控的基本路径。

（一）厘清致命性自主武器系统军控的相关概念

推进致命性自主武器系统军控，首先必须将致命性自主武器系统的相关概念界定清晰，否则很难实际操作。目前来看，厘清以下几个关键概念和问题尤为重要。

第一，如何理解“自主性”？笔者认为，保罗·沙瑞尔的观点颇有参考价值，能够为我们理解自主性提供较为清晰的框架。他认为，理解自主性主要有三个维度，分别是机器承担的任务类型、执行任务时的人机关系以及机器决策的智能程度。从任务类型来看，能够对机器的自主性进行区分，比如有人驾驶汽车上的自动刹车系统，从驾驶汽车这个任务来看它拥有一定自主性但又非完全自主，但从刹车角度来看则是完全自主的。从人机关系（人在观察—判断—决策—行动 OODA 回路中的作用）这一维度来看，自主武器可以分为“人在回路内”（human in the loop），也称为半自主武器；“人在回路上”（human on the loop），也称为有监督的自主武器；以及“人在回路外”（human out of the loop），也称为完全自主武器。在半自主系统中，机器执行一部分任务之后，会停下来等待人类的指令授权再采取下一步行动；在有监督的自主武器系统中，机器能自主进行观察、判断、决策和行动，但人可

以监督系统运行并在必要时进行干预；而完全自主武器系统则能够独立完成观察、判断、决策、行动这一回路。按照机器的智能程度，可以将自主系统划分为自动的（automatic）、自动化的（automated）和自主的（autonomous）。自动系统是指简单的、基于阈值的系统，很少涉及决策过程，如老式的恒温器。自动化系统则是指一种相较自动系统更为复杂的基于规则的系统，需要考虑更多的输入条件并权衡变量，例如现代数字化可编程恒温器。自主系统的智能程度更高，是目标导向自我指导以及内部机制难以被用户掌握的复杂系统。[①] 此外，在界定自主性的概念后，还需要对自主武器的自主程度进行量化衡量，并根据不同程度制定相应规则。如，美国海军研究办公室及空军研究实验室（AFRL）就定义了从遥控到完全自主集群的 10 个自主控制级别（Autonomous Control Level, ACL），[②] 对无人机的自主程度进行了相对量化的衡量，这一尝试值得国际社会参考和借鉴。

第二，什么是致命性自主武器系统？要禁止哪一类致命性自主武器？是禁止其研发、生产和使用等全部环节，还是只禁止其中的一个或几个环节？如果国际上没有普遍接受的定义能够界定清楚究竟什么是自主武器，对于该武器系统的军备控制也就无从下手。关于致命性自主武器系统的概念，存在多种说法。综合来看，主要是指不需要人类干预，能独立选择和打击目标并能致人丧生的武器系统。[③] 红十字国际委员会对该概念的界定值得重点参考，即“在没有人力干预下能够选择（寻找、察觉、确认、追踪、选择）和攻击（使用武力、压制、损坏、摧毁）目标的武器系统”，自主性是其核心功能。[④] 这一定义相对客观清晰，指出了致命性自主武器系统的致命性、自主性等关键特征。

① [美]保罗•沙瑞尔：《无人军队：自主武器与未来战争》，第 32—37 页。

② 参见：Bruce T. Clough, “Metrics, Schmetrics! How The Heck Do You Determine A UAV’s Autonomy Anyway?” 2002, https://pdfs.semanticscholar.org/0025/3d204cffb9ab60663f79bbc0a68e014e9ede.pdf?_ga=2.30172151.918544268.1579618107-323294172.1579618107。

③ Rebecca Crootof, “The Killer Robots Are Here: Legal and Policy Implications,” *Cardozo Law Review*, Vol. 36, No. 5, June 2015, http://cardozolawreview.com/wp-content/uploads/2018/08/CROOTOF.36.5.pdf.

④ International Committee of the Red Cross, “Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?” March 29, 2018, https://www.unog.ch/80256EDD006B8954/(httpAssets)/42010361723DC854C1258264005C3A7D/$file/CCW_GGE.1_2018_WP.5+ICRC+final.pdf, p.4.

第三，人类对武力使用应当保持何种程度的判断和控制？目前，几乎所有参与联合国《特定常规武器公约》探讨致命性自主武器系统问题的与会方都强调必须保持人对武力使用的控制，保持人类控制的概念一定程度上成为关于如何处理杀手机器人这一辩论的核心，[①] 但是对于保持人类对武力使用何种程度的判断和控制这一问题存在不同意见。笔者认为，至少要确保在武器系统上进行最低限度的人类判断和控制，即在确定武力行动是否合法、是否进行攸关生死的攻击行动等问题上必须有人类的干预和参与，决不能放任机器自主做出此类决策。不然，就会使得武力使用脱离人类的控制，引发后续一系列问责和失控问题，甚至会危害人类的生存。而在不涉及人员伤亡的领域中（如卫星数据分析和核生化环境勘察、救援等），可以给机器更多的自主空间，以提升作战效率。

（二）明确针对致命性自主武器系统的核心原则

目前，国际人道法（也称国际人道主义法或武装冲突法）是国际上维护和平、限制战争危害人道的重要保障，其中的主要原则也为国际社会广泛认可。笔者认为，现有以国际人道法和国际人权法为主体构成的法律框架依旧适用于对致命性自主武器系统的约束和管控。要推动致命性自主武器军控进程，保护基本人权，需要明确国际人道法主要原则在这一领域的适用性。

第一，遵守国际人道法中的区分原则。《日内瓦公约第一附加议定书》第 48、51、52 条规定了区分原则，[②] 明确冲突各方无论何时均应在平民和战斗员之间以及在民用物体和军事目标之间加以区别，冲突一方的军事行动仅应以军事目标为对象，即禁止或限制使用某些可被认为具有滥杀、滥伤作用的武器系统。[③] 由此可知，使用致命性自主武器系统直接攻击平民、民用

① 据笔者了解，目前主要是这一句式：“____人类____”。关于“人类”之前的表述包括“有意义的”（meaningful）、“适当的”（appropriate）、“必要的”（necessary）和“重要的”（significant）等，“人类“之后的表述包括“控制”（control）、“介入”（involvement）和“干预”（intervention）等。具体使用哪一个表述，各方存在明显争议，目前尚未统一。

② 《日内瓦公约第一附加议定书》全称为《1949 年 8 月 12 日日内瓦四公约关于保护国际性武装冲突受难者的附加议定书》。

③ “1977 Additional Protocol I to the Geneva Conventions,” Columbia Law School, June 8, 1977, https://web.law.columbia.edu/sites/default/files/microsites/gender-sexuality/Protocol%20I%20and%20II.pdf.

物体（不论是故意还是失误），无疑都是违反国际人道法的。在这一点上，致命性自主武器系统这一新兴武器与传统武器没有太大区别，国际人道法只从区分的结果角度规定平民和民用物体不得受到攻击和恐吓，并未规定区分的过程应当由谁来完成。因此、无论是由武器系统操作员对目标进行区分，还是由武器本身通过预先设定程序进行自主识别，区分的方法都须遵守这一规定，不能模糊了军民的界限。此外，第一附加议定书第 51 条第 4 款第 1 项还规定禁止“不分青红皂白的攻击”，即不能进行不以特定军事目标为对象的攻击。据此，如果致命性自主武器系统发动未区分军事目标与民用物体的攻击，即构成违法。因此，明确强调致命性自主武器及其应用必须遵守区分原则，才能保护平民，维护人道主义法和基本人权。

第二，要遵守国际人道法中的比例原则。国际人道法关于比例原则的规定，主要体现在《日内瓦公约第一附加议定书》第 51 条第 5 款第 2 项以及第 57 条第 2 款第 1 项。这两条规定禁止或限制使用某些可被认为具有过分伤害力的武器系统。[①] 它们涉及了“过分”的判断问题，其中明确指出，判断是否“过分”，不是简单地将平民伤亡与敌方伤亡进行比较，而是应当根据具体情况判断发动攻击是否真正符合军事必要这一原则，以确定可接受范围内的附带伤害。作为国际人道法基本原则之一，比例原则旨在禁止发动可能附带使平民生命受损失、平民受伤害、平民物体受损害并且与预期的具体和直接军事利益相比损害过分的攻击。一般而言，致命性自主武器系统的自主程度很高，能够根据预设的程序观察、分析、瞄准和打击目标。因此致命性自主武器系统能否在无人操作或监控的情况下确定具体攻击任务是否具备相应的军事必要性，是判断其攻击合法性的关键。遵守比例原则，意味着在某次特定的军事行动中所获得的军事利益，应该超过其所造成的附带性平民损伤。这就要求在攻击前，武装冲突的一方必须将预期平民伤亡与行动的预期军事效果予以衡量，需要在既考虑总体环境因素又顾及具体因素的情况下，对一次攻击是否遵守比例原则的要求进行主观的个案评估。对于依赖人工智能算法的致命性自主武器系统而言，这种既考虑战场环境客观因素又顾

① “1977 Additional Protocol I to the Geneva Conventions.”

及社会文化主观因素的自主个案评估将是十分困难的任务。有鉴于此，必须明确致命性自主武器系统遵守比例原则，如果该系统不能达到这一要求，便不能允许其自由发展、生产和使用。

第三，要符合国际人道法中“攻击中的预防措施”（Precautions in Attack）原则以及“退出战斗”（Hor De Combat）法则。其中，“攻击中的预防措施”原则强调准备攻击或决定攻击的人要“采取一切可行的预防措施”，以避免伤害平民。[①] 与比例原则相似，这一要求对于致命性自主武器系统的使用而言，要求至少要保证人类指挥官或武器系统操作员对于攻击行为进行恰当的判断，对军事必要性和平民附带损伤风险进行慎重权衡，不可任由机器做出攸关生死的决策，造成不必要的平民伤亡而无法追责。“退出战斗”法则也是国际人道法中的重要规则，其禁止攻击“已经投降或无法战斗”的战斗人员。具体而言，“退出战斗”指的是被俘虏、明确表示投降、无意识或因伤病而丧失作战能力无法进行自卫的参战人员。[②] 在致命性自主武器系统这一问题上，目前的技术发展还不足以使机器系统辨析敌方的假投降和假丧失作战能力等“诈骗”和“愚弄”行为，很难使得这类武器系统符合这一原则。在此背景下，可以考虑禁止研发和使用以人为目标的致命性自主武器系统，而对以物体为目标的自主武器系统可以相对宽容。

第四，遵守国际人道法中的公众良知原则。致命性自主武器系统不仅需要遵守国际人道法中规定的以上具体原则，还需在上述原则未顾及的方面，根据符合反映人道主义关怀的“马顿斯条款”（Martens Clause），对致命性自主武器系统的使用进行有效规制。具体而言，国际人道法中的“马顿斯条款”规定：“凡遇有本条文中未规定之事项，则有种种国际法之原则，从文明人民之惯例上，从人道之原则上，自良心之要求上，发生事变之两交战国与其人民之间，应在此原则之保护与支配下。”[③] 这意味着“马顿斯条款”

① 转引自[美]保罗•沙瑞尔：《无人军队：自主武器与未来战争》，第 291 页。原文参见《日内瓦公约》第 57 条：Article 57(2)(a)(ii), Protocol Additional to the Geneva Conventions of August 12, 1949, (Protocol I); and “Rule 15: Precautions in Attack,” Customary IHL。

② 同上，第 291—292 页。原文参见《日内瓦公约》第 57 条：Article 41, Safeguard of an Enemy Hors de Combat,” Protocol Additional to the Geneva Conventions of 12 August 1949 (Protocol I)。

③ Protocol Additional to the Geneva Conventions of August 12, 1949, and relating to the

禁止有违“公众良知”指引的武器系统。致命性自主武器系统将人类决策的责任转移给旨在夺取人类生命的武器系统，构成了对国际人道法的根本冲击。致命性自主武器系统的自主性正在不断提升，倘若不加以有效管控，可能会接管战斗人员所有的决策权。是否应该赋予机器人以生杀予夺的权力？对于这一问题现有国际法并没有相应的明确条文规定，但是从国际人道法精神来看，致命性自主武器系统对人的生命的伤害或剥夺，应符合国际人道法隐含的公众良知要求。换言之，人工智能技术足够成熟，致命性自主武器系统背后所代表的“致命决策去人类化”要具有人道主义上的正当性，最根本的问题仍在于如何守住“马顿斯条款”所诠释的人道底线，这将是人类不会妥协也不应妥协的一个问题。面对致命性自主武器系统可能对人的生命及尊严带来无法预知的威胁和损害，国际法和军控机制的完善应尽快弥补这一法律制度的漏洞，守住人道主义与道德底线。同时需要指出的是，公众良知这一原则也存在一些显著问题。[①] 比如，公众良知与公众舆论能画等号吗？是哪个国家的公众良知？（美国公众的良知？中国公众的良知？全人类的良知？）如何才能对公众在致命性自主武器这一问题上的良知进行有效测量？这些问题还有待国际社会的进一步探讨和解决。

（三）建立和完善致命性自主武器系统军控的相关机制

世界各国为了遏制战争进行了长期不懈的努力，自 1856 年《巴黎海战宣言》（Paris Declaration on Naval War）到 2013 年联合国《武器贸易公约》（Arms Trade Treaty, ATT），国际社会逐步建构起相对完备而有效的国际法和军控机制。致命性自主武器系统军控同样需要织牢制度之网，在遵循现有国际法和军控机制的基础上，稳步推进这一领域的国际安全合作。

第一，完善《日内瓦公约第一附加议定书》第 36 条。对于致命性自主武器系统这一新式武器的出现，除须遵守《海牙公约》和 1949 年《日内瓦公约》及其 1977 年两个《附加议定书》等法律制度外，还需要受 1977 年《第

Protection of Victims of Non-International Armed Conflicts (Protocol II), June 8, 1977, https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/7c4d08d9b287a42141256739003e636b/d67c3971bcff1c10c125641e0052b545.

① 关于公众良知原则在自主武器这一领域的适用问题的详细探讨，可参见[美]保罗•沙瑞尔：《无人军队：自主武器与未来战争》，第 297—299 页。

一附加议定书》第 36 条和《特定常规武器公约》及其附加议定书的直接规制。尤其是 1977 年《第一附加议定书》第 36 条实施措施《新武器、作战手段和方法法律审查指南》指出，“在研究、发展、取得或采用新的武器、作战手段或方法时，缔约一方有义务断定在某些或所有情况下，该新武器、作战手段或方法的使用是否为本议定书或适用于该缔约一方的任何其他国际法规则所禁止。”① 这一规定旨在通过在发展或取得武器之前断定其合法性的办法，来防止使用那些在一切情况下均违反国际法的武器，并对那些在某些情况下违反国际法的武器的使用加以限制。但从目前建立武器合法性国内审查机制国家寥寥无几的情况来看，这一机制还存在诸多明显缺陷，例如，缺乏交流与互助机制等。有鉴于此，国际社会应正视这些问题，并积极推动这一机制的补充完善和有效落实。

第二，以补充议定书形式将致命性自主武器系统军控纳入《特定常规武器公约》框架。历史上，联合国《特定常规武器公约》对于杀伤人员地雷、燃烧武器、激光致盲武器等对平民可能带来过分杀伤和不必要痛苦的常规武器进行了相对有效的军备控制。致命性自主武器系统作为一种新出现的特定常规武器，国际社会要求遵照国际法和军控机制的基本精神，将其纳入《特定常规武器公约》框架加以有效规制的呼声也在与日俱增。在此背景下，可考虑借鉴历史经验，以《特定常规武器公约》补充议定书的方式，制定对完全自主的致命性武器系统进行规范和限制的措施。同时，在致命性自主武器系统军控因其应用的实际影响尚有诸多争议的情况下，从相对较易达成共识的增强武器审查透明度和国际监管适时介入等方面着手有利于形成共识，将相关军控努力引向以补充议定书形式纳入《特定常规武器公约》框架。

第三，对完全自主的致命性武器系统实行监管下的预先禁止。对致命性自主武器系统实行军控的根本依据，主要来自其脱离了人的控制后可能会出现误杀误伤、滥杀滥伤以及过分伤害这一考量。而脱离人类控制的完全自主的致命性武器系统无论是从法律、伦理还是战略稳定层面来看，都不符合人

① “1977 Additional Protocol I to the Geneva Conventions,” Columbia Law School, https://web.law.columbia.edu/sites/default/files/microsites/gender-sexuality/Protocol%20I%20and%20II.pdf.

类整体价值和各国利益。有鉴于此，可以考虑预先禁止完全自主的致命性武器系统，而不是禁止遥控作战兵器、巡航导弹等低自主性武器。这一军控成功的关键在于建构必要的监管框架，确保对战争中自动化程度越来越高的武器进行区别，然后视其情况进行控制。通过对这一武器系统进行强有力的监管，确保其在选择和攻击目标时受人类的控制，保证它作为一款致命性武器对人的攻击背后总可以找到对应的人来承担责任。针对那种完全自主的、排除了任何人类控制的武器，才应禁止其研发、部署和使用。在这种设想的情景下，人至少仍会是武器系统使用“回路”的一部分，不论致命性武器系统的自主程度如何，都必须确保武器系统在选择和攻击目标时还是受人控制的，人要对其造成的后果负责。换言之，致命性自主武器军控机制应是一个建立在有效监管框架下的甄别性禁止，对于没有人类任何控制就能完全自主致命的武器实行禁止，而非禁止所有的自主武器。

第四，建立具有约束力的致命性自主武器系统军控核查机制。致命性自主武器系统军控机制的重要功能是承担其问责使命。如果仅仅将滥杀、滥伤归之于“由于系统错误造成问题”，可能会出现无辜平民被杀戮但没有任何一方需要对此承担责任的窘境。针对致命性自主武器系统容易引发的问责空白问题，应该构建一套国际监管框架下的甄别禁止机制来为不同自主程度的军事机器行为确定责任。责任划分应该跟随指挥链，需要明确使用不同自主程度致命性武器系统的组织或个人应该对其行动负责；“系统故障”不能成为不必要伤亡的正当理由。有关研发者应该清楚地评估“系统故障”，并预告使用者，使用者就应在权衡“系统故障”的基础上，对不同自主程度致命性武器系统可能出现的错误和罪行负责。如果将这一理念和思路纳入国际人道主义法和国际人权法——目前只管辖人类行动者而非机器，那么这些国际法可能将为致命性自主武器系统的监管提供充分依据。对于非完全自主的致命性武器，因有“人”的控制因素在指挥链中，那么，相应的“人”不管是研发者、生产厂商，还是使用者就必然要承担相应的责任。因此，问题的关键就落到了国际社会到底能不能发现致命性自主武器系统杀害无辜平民的事件，能不能在核查证据的基础上坚决将犯罪分子绳之以法，这就要求对致

命性自主武器系统军控建立有效的核查机制。

四、中国参与致命性自主武器系统军控的策略

当前，中国已经认识到致命性自主武器系统军控的重要性，并已积极参与到这一进程中。[①] 对此有以下几种策略可供参考。

（一）联合国际社会推动联合国框架下的致命性自主武器系统军控

推动致命性自主武器系统军控是大势所趋，是有关人类共同安全的时代命题。各国需要通过一个成熟的平台共同探讨和解决这一问题，制定务实可行的方案，以应对其发展可能带来的挑战和风险，而联合国则是一个很好的平台。在联合国框架下推动致命性自主武器系统军控，既拥有良好的道义支持和历史基础，也更符合中国的国家利益，有助于把握规则制定主动权。从现有的国际法和军控机制体系来看，《特定常规武器公约》作为常规军控领域的重要法律框架之一，其宗旨和原则得到越来越多国家的认同，该机制也已成为致命性自主武器系统军控的核心平台。自 2014 年以来，《特定常规武器公约》机制已召开多次会议，就致命性自主武器系统的概念、特点、技术、伦理、法律、军事等问题展开广泛讨论。为此，中国作为《特定常规武器公约》缔约国之一，应表明对致命性自主武器系统滥用引发人道主义危机的高度关切，积极参与和推动此框架下的致命性自主武器系统军控进程，宜主张在平衡处理人道主义关切和各国正当军事安全需要的基础上，不断完善相关国际法律机制，支持致命性自主武器系统军控以补充议定书形式纳入《特定常规武器公约》框架。此外，还可建议通过联合国决议或者国际条约在联合国设立国际人工智能中心，由各国政府代表、技术专家以及相关非政府组织代表组成，致力于推动致命性自主武器系统的管理以及国际规则制定。技术专家可以专注于解决人工智能的算法歧视、黑客攻击等问题；安全和法律专

① 例如，中国作为联合国《特定常规武器公约》缔约国，已经多次参加了关于致命性自主武器系统政府专家组会议。参见："Position Paper Submitted by China" (CCW/GGE.1/2018/WP.7), April 11, 2018, https://www.unog.ch/80256EDD006B8954/(httpAssets)/E42AE83BDB3525D0C125826C0040B262/$file/CCW_GGE.1_2018_WP.7.pdf。

家则可致力于制定相关的法律和规则。此外，致命性自主武器系统的军控还可以尝试从容易达成共识的方面入手，比如先推动禁止严重威胁国际安全和战略稳定性的核自主武器的研发、生产和使用，然后逐渐进入更艰难的部分。

（二）明确本国立场，强调自身关于致命性自主武器系统的核心观点

第一，强调致命性自主武器系统应该被理解为完全自主的致命武器系统。致命性自主武器系统的定义如果界定模糊，或者包含半自主和自动化武器系统，那么推动致命性武器系统的军控将面临重重阻碍，也无益于世界和平与发展。因为，目前很多武器系统都或多或少带有一定的自动化特征（如精确制导武器），如果将这些也视为致命性自主武器系统进行军备控制，非但不现实，也会损害本国利益。此外，要明确致命性自主武器系统的主要特征，以下几点尤为重要。一是致命性，即自主武器系统具有致人伤亡的功能；二是自主性，即武器系统在执行任务的整个过程中不需要人为干预和控制；第三，不可终止性，即一旦启动就无法终止该武器系统的攻击行为。[①] 将致命性自主武器系统界定为完全自主的致命自主系统，既能表明制止“杀手机器人”的态度，也能在推动军备控制时不涵盖现有的一些高自动化或半自主武器系统，这样面临的阻力就会相对小一些，也符合本国利益。

第二，表明坚决不能将攸关人类生死存亡的决定权让渡给机器的态度，要保持人对武器系统和武力使用的有效控制。当前，各国在保持人类对武力使用的控制上基本达成共识，这是攸关人类尊严和生死存亡的道德底线和红线。一旦越过这条红线，杀人机器人就可能泛滥并挣脱人类的控制和束缚，整个人类社会就可能面临灭顶之灾。美国《国防部指令 3000.09》也指出：“自主武器系统的设计应允许指挥官和操作人员对使用武力行使适当程度的人类判断。”[②] 其他很多国家也意识到了这一问题的重要性。在这一点上，中国也要坚守底线，反对使用完全的自主武器，坚持人类要在自主武器系统研发、生产、部署的整个过程中发挥基本的控制作用。

第三，强调既要看到致命性自主武器系统对于军事领域可能带来的积极

① “Position Paper Submitted by China” (CCW/GGE.1/2018/WP.7), p. 2.

② U.S. Department of Defense, *Directive 3000.09, Autonomy in Weapon Systems*, 2012, p. 2, https://fas.org/irp/doddir/dod/d3000_09.pdf.

影响，也要防范其危险。自主武器系统在很多领域能发挥积极作用，比如可以增强对平民和民用目标的军事感知能力，高度适用于面临核武器、生物武器和化学武器威胁的环境。与此同时，致命性自主武器系统的开发和使用将降低战争的门槛和使用国的战争成本。此外，目前来看，致命性自主武器系统还无法有效区分作战人员和平民，容易造成滥杀、滥伤。有鉴于此，要呼吁所有国家采取有效的预防措施，避免致命性自主武器系统伤害平民。

第四，还要明确致命性自主武器系统军控不应当妨碍民用领域的人工智能技术创新和应用。当前，人工智能正成为推动新一轮科技、产业和社会变革的核心驱动技术，为提升中国人民的生活水平以及军队实力创造了机遇。习近平主席也强调，“加快发展新一代人工智能是我们赢得全球科技竞争主动权的重要战略抓手，是推动我国科技跨越发展、产业优化升级、生产力整体跃升的重要战略资源。”① 尽管人工智能等新兴技术是致命性自主武器的基础技术，但它们已经在许多国家的经济和社会发展中得到广泛应用。世界各国、各界需要客观而充分讨论人工智能技术的影响，不能草率预设前提或预判结果，因为这可能会阻碍人工智能技术发展。在此背景下，中国应当呼吁禁止致命性自主武器系统的使用，而不是禁止其支撑技术的研发和生产。倘若在还未完全了解致命性自主武器系统的情况下，就草率禁止相关研发，无异于作茧自缚，将会制约中国人工智能技术的正常发展和合理运用，反而给恐怖分子和无视国际规范的国家以可乘之机。

（三）确立底线思维，做好致命性自主武器系统军控受挫的相关准备

自特朗普执政以来，美国相继退出《跨太平洋伙伴关系协定》《巴黎协定》、联合国教科文组织、《伊核协议》等多个国际协议和组织，令国际社会咋舌。更令人失望的是，2019 年 8 月 2 日，美国正式退出《中导条约》。② 美国此举无疑会加剧大国军备竞赛和地缘政治博弈，侵蚀国际核不扩散体系，

① 习近平:《推动我国新一代人工智能健康发展》,习近平在中共中央政治局第九次集体学习的讲话，人民网，2018 年 10 月 31 日，http://cpc.people.com.cn/n1/2018/1031/c64094-30374719.html。

② 《中导条约》的全称为《苏联和美国消除两国中程和中短程导弹条约》。1987 年 12 月 8 日，美苏在华盛顿签署了这一条约，规定两国不允许研制和使用射程在 500—5500 公里的弹道导弹以及陆基巡航导弹。

增加国际军控难度，影响世界对于其他军控协议有效性的信心。

历史表明，军备控制需要大国间的协调一致才能获得实质性推进。但是，美国在推进致命性自主武器系统军控方面并不积极，反而在不遗余力地推动人工智能的军事化。具体而言，特朗普签署了《维护美国在人工智能时代的领导地位行政命令》（Executive Order on Maintaining American Leadership in Artificial Intelligence）；美国国防部推出“第三次抵消战略”，将人工智能作为核心技术，相继发布《美国陆军无人机系统路线图（2010—2035）》和陆军《机器人与自主系统战略》，并在 2018 年出台人工智能战略，成立“联合人工智能中心”和“算法战跨职能小组”，全力推动人工智能军事化；美国国防部高级研究计划局（DARPA）也相继推出“小精灵”“指挥官虚拟参谋”等多个与人工智能相关的研发项目。[①] 在此背景下，致命性自主武器军控的前景并不乐观，在这个领域很难形成一份具有法律约束力的国际文件。即使美国同意签署一份致命性自主武器系统的军控协议，也未必会切实遵守，更难预料其在何时又会像退出《中导条约》一样退出这个协议。因此，中国不能自缚手脚，而应做好两手准备，在积极推动致命性自主武器系统军控的同时，也要推进人工智能的发展及在国防领域的合理运用。

结 束 语

致命性自主武器系统军控前景可期，但依然任重道远。从军备控制的历史经验看，能够取得成功的案例基本要符合以下两个条件或至少其中之一：一是该武器系统威胁国际安全和战略稳定，如核武器和反导系统；二是该武器系统严重挑战人道主义和人权，对平民构成极大威胁，如生化武器、激光致盲武器等。从现有情况看，致命性自主武器系统对这两大领域都可能产生

① 参见 White House, *Executive Order on Maintaining American Leadership in Artificial Intelligence*, February 2019, https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence; U.S. Department of Defense, *Summary of The 2018 Department Of Defense Artificial Intelligence Strategy*, February 2019, https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF; 龙坤、朱启超：《算法战争的概念、特点与影响》，《国防科技》2017 年第 6 期，第 36—42 页。

很大挑战，但具体影响方式目前仍不明晰。可以预见，人工智能军事化已成为不可阻挡的趋势，当前能够做的只是限制其发展领域、程度，为其划定红线，避免自主武器破坏战略稳定和威胁平民安全。

致命性自主武器系统军控攸关个人安全、国家利益和人类前途命运，作为联合国安理会五个常任理事国之一和最大的发展中国家，中国应予以高度关注，积极参与国际社会对致命性自主武器系统军控的探讨，大力推动联合国主导军控机制建构的国际谈判。在推进致命性自主武器系统军控的进程中，中国应承担起应有的大国责任，在维护本国国家利益的同时，推进人类命运共同体建设，守护人类安全和福祉。一方面，中国应明确自身的核心观点和原则，支持以补充议定书形式将致命性自主武器系统军控纳入《特定常规武器公约》框架，构建对致命性自主武器系统有效监管和核查的军控机制。另一方面，中国也决不能自缚手脚，在积极参与和推进联合国主导的致命性自主武器系统军控机制构建的同时，也要加强人工智能发展的战略规划和总体布局，夯实支撑人工智能发展的工业基础和技术支撑，推动军、民用资源的统筹利用，加快推进中国特色军民融合式人工智能的发展进程，这将有利于提升中国在相关领域的实力以及在国际致命性自主武器系统军控谈判过程中的话语权。

[责任编辑：樊文光]