The Ethics of Intelligent War
Risks and Countermeasures

智能化战争伦理
风险与应对

张煌 王伟 著

国防科技大学出版社
National University of Defense Technology Press

人工智能的军事化发展趋势不可逆转，智能化战争的脚步已经迫近。人工智能究竟会以怎样的方式影响我们身处的安全环境和未来的安全前景？它会带来更人道的作战方式和不流血的战争，还是如同开启潘多拉魔盒一般，将人类社会彻底拖入恐怖的终结时代？



上架建议：军事

ISBN 978-7-5673-0600-4

9 787567 306004 >

定价：78.00元

责任编辑：周　蓉

封面设计：陈芷怡

国防科技大学出版社
官方微信公众号

The Ethics of Intelligent War
Risks and Countermeasures

智能化战争伦理
风险与应对

张煌 王伟 著

国防科技大学出版社
·长沙·

图书在版编目（CIP）数据

智能化战争伦理：风险与应对/张煌，王伟著. —长沙：国防科技大学出版社，2023.11

ISBN 978-7-5673-0600-4

Ⅰ.①智… Ⅱ.①张… ②王… Ⅲ.①高技术战争—伦理学—研究 Ⅳ.①E866-05

中国国家版本馆CIP数据核字（2023）第059962号

智能化战争伦理：风险与应对

ZHINENGHUA ZHANZHENG LUNLI：FENGXIAN YU YINGDUI

张 煌 王 伟 著

责任编辑：周 蓉

责任校对：朱哲婧

出版发行：国防科技大学出版社　　地　址：长沙市开福区德雅路109号

邮政编码：410073　　电　话：（0731）87028022

印　制：国防科技大学印刷厂　　开　本：710×1000　1/16

印　张：16.75　　字　数：210千字

版　次：2023年11月第1版　　印　次：2023年11月第1次

印　数：1-1500册

书　号：ISBN 978-7-5673-0600-4

定　价：78.00元

版权所有　侵权必究

告读者：如发现本书有印装质量问题，请与出版社联系。

网址：https://www.nudt.edu.cn/press/

序

学术新星何以升起

不久前，国防科技大学年轻的副教授张煌博士被遴选为湖南省湖湘青年英才，成为中国学术界一颗冉冉升起的新星。在十余年的治学之路上，张煌已给世人留下了一连串深刻的印记，尤其是近年在战争伦理、军事战略与国防科技发展战略的交叉领域，发表了一系列见解新颖、论证严谨、功力非凡的长篇鸿文，赢得了学界普遍的赞誉，学术影响力与日俱增，如日之升。

多年以来，我与张煌既是师生，又是学友。作为其潜心一志治学的见证人，深感在研究生大规模扩张、青年学子源源不断涌入学术领域，而学术浅泛、雷同等问题不容小觑的当今学界，张煌艰辛修炼、追求卓越、坚守初心的治学品格，当给怀抱学术理想的年轻学人宝贵的启迪。

一、深厚的学术情怀

无论在读研究生期间，还是留校工作后，张煌总是脚踏实地，立足岗位，胸怀远大学术理想，不受眼前功利的诱惑，不作好高骛远的空想，不求易于消逝的时髦，不掷宝贵年华于物质享受。我记得，在硕士、博士阶段，张煌谨记钱学森先生关于大学及后续五年是打好学术基础关键时期的教导，专心致志学习，从未谈过恋爱，而与文香四溢的好书为伴。读书、思考、讨论、写作，是张煌早在求学时期就形成的一种生活方式，直至现在也仍然是张煌生活的重心。这种深厚的学术情怀源远流长，是古今中外一切有成就的学者强大而持久的学术动力。

数学大师华罗庚，只有初中学历，自学成才，“研究第一，学问至上”的学术情怀，使他放弃了攻读剑桥大学博士学位的机会。1936 年，华罗庚获得中华文化教育基金会的资助，赴剑桥大学深造，导师哈代是当时叱咤风云的国际数学大师，十分欣赏华罗庚的才华，留了一张纸条给秘书海尔布伦：“华来时，请转告他，他可以在两年之内获得博士学位。”海尔布伦问华罗庚：“你打算攻读哪一门课程？我们将给你帮助。”不料华罗庚竟回答说：“谢谢你的好意，我只有两年的研究时间，自然要多学点东西，多写些有意思的文章，念博士不免有些繁文缛节，太浪费时间了，我不想念博士学位，我只要求做一个访问学者。我来剑桥大学是为了求学问，不是为了学位。”海尔布伦感到意外地说：“东方来的人，不稀罕剑桥大学的博士学位者，你还是第一个，我们欢迎

你这样的访问学者。”[①] 华罗庚只做学问、不攻读博士学位的做法，未必适合每个出国留学者，但是唯求真才实学的纯粹学术情怀如此强烈确实令人惊赞。正因有这样的学术情怀，华罗庚在留学剑桥短短两年里，发表了十余篇高质量论文，“特别是他解决了完整三角和估计这一历史名题，……导致了对华林问题的重要改造”。[②] 28岁的华罗庚由此跻身世界数论大师的行列。

杰出的文学大师钱钟书，是公认的奇才，因纯粹深厚的学术情怀，留学牛津大学时，不为名位所动，放弃了攻读牛津大学博士学位的机会，谢绝了牛津大学的中文讲师礼聘，并婉拒了与来访的美国纽约大学现代文学教授合作撰写文学批评史的倡议。[③] 抗战结束后的1946年，曾有人推荐钱钟书到联合国教科文组织任职，许多知识分子可能视此为难得的好机会，然而出于“贡献一生，做做学问”的学术情怀，钱钟书谢绝了这份好意。失之东隅，收之桑榆。深厚纯粹的学术情怀，使钱钟书在中国古代文学与中西文学比较研究方面留下了《谈艺录》与《管锥篇》两大经典名著。

二、独特的研究方向

在学者成长过程中，通常学术情怀的养成要早于研究方向的确立，这是与人才的立志一般要早于学问的修炼这一普遍规律相一致的。杨振宁至迟在大学时代已形成强烈的学术情怀，却迟迟

① 王元. 华罗庚 [M]. 北京：开明出版社，1994：66.

② 王元，杨德庄. 华罗庚的数学生涯 [M]. 北京：科学出版社，2005：2.

③ 孔庆茂. 钱钟书传 [M]. 南京：江苏文艺出版社，1992：77－78.

没有找到适合自己的研究方向，直到赴美攻读博士学位时，明显显露出长于理论而拙于实验的个性特征后，才在导师泰勒的建议下专攻理论物理。这一研究方向的确立奠定了杨振宁世界顶尖物理学大师的基础。1969 年诺贝尔生理学及医学奖得主德尔布吕克，则在 1930 年获哥廷根大学哲学博士学位（导师玻恩）后，在物理学大师玻尔的指点下，审时度势，毅然离开已近完美的现代物理学领域，而转向方兴未艾的生命科学世界，确立起前途无量的新研究方向，并因发现病毒的复制机制和遗传结构而荣获诺贝尔奖。更值得赞赏的是，汉字激光照排的发明者王选院士，从科班的计算数学专业出发，经计算机研发实践的感悟和计算机发展史文献的阅读与思考，独自发现要做一流的计算机研究，必须走硬件与软件相结合的方向与路径，正是沿着这一研究方向，加上扎实的数学功底与出色的英语水平，攻克了汉字激光照排这一世界印刷技术的制高点。

一般说来，研究方向是指在广阔的学术领域中一个不大不小的独特范围，可在导师指导下确定，但更多的是由学者自身敏感性而确定。这种确定虽有太多只可意会不可言传的神秘性，但依然有某种有章可循的理性可以启迪后学。总的说来，确定一个研究方向，需要考虑自身的兴趣、优长及知识结构。杨振宁招收研究生时，要反复询问两个问题：是否对物理学真有兴趣？是否擅长此道？外在环境与条件能否支持这样的研究方向，也是考虑的重要因素。青年学子为确立切实可行而前景可观的研究方向，不仅要深于理性思考，更要敏于实践的指引。上述举例中，大师玻尔的理性思考直接指引了德尔布吕克，而杨振宁与王选的选择，

更多受益于实践和引导。

本书作者艰辛探索并睿智选择研究方向的过程，可说是体现上述规则的一个典范。自读研开始，他便一直关注军事与科技交叉的国防科技问题及其相关领域，曾以国防科技人才的成长培养为题完成博士论文。在国防科技大学国家安全与军事战略研究中心工作期间，兴趣转向国防科技与军事战略研究，到新加坡国立大学的访学，又使学术视野进一步拓展。经多年的积累、思考与实践，终于选定国防科技、战争伦理、军事战略之间的交叉领域为研究方向，这是一个充满空白而对强军兴国极端重要的学术领域。经潜心深耕，他取得了一系列重要成果，在这一研究方向的前沿阵地已占有一席之地，深得同行与领导的赞赏。

三、长期的坚持不懈

当年曾国藩的四大弟子之一而后成中国近代外交家的黎庶昌，在33岁时致信曾国藩，为自己当无建树、默默无闻而苦恼。曾国藩在回信中写道："古来如颜子立德，周郎立功，贾生立言，均在少壮。然千古曾有几人？其余贤哲代兴，树立宏大，大抵皆在四十以后耳。以仲尼之圣而不惑亦待四十。"[①] 当代世界信息大爆炸，知识激增，治学者学习训练的时间更为漫长，有志科研的青年在闻达以前的学术寂寞期必更为久长。一个才华出众而意志力薄弱的人，最终结果常常不及中等才具而意志力坚强的人，这在人类文化史上屡见不鲜。年轻的治学者如果对所选的研究方向有

① 唐浩明. 曾国藩书信［M］. 长沙：岳麓书社，2016：235.

独特的敏感性，又有耐得住寂寞、坚韧不拔的意志力，则有可能在40岁以前突破瓶颈，脱颖而出。本书后记中，作者回顾了自己的治学道路，坦陈应对繁重工作和战胜孤独彷徨的心路历程。正是依靠坚持不懈的非凡毅力，作者度过了学术之路上最易动摇放弃的脱茧蝶变期，并在40岁之前开辟出一片学术领地，跻身放射独特光芒的学术新星之列。

关于作者本书的具体研究，我想说，在新军事技术与战争伦理交叉领域的前沿探索，恰如科学技术的前沿探索一样，是一种充满挑战而永无止境的崇高事业，能在年轻时代，勇于深入这一交叉学科前沿的漩涡中心，参与这样重大而前景广阔的学术研究，是作者本人的幸运，也是战争伦理学与科技伦理学的幸运。

朱亚宗

2023年9月

前 言

人工智能的发展及其军事应用，给予这个时代太多的不确定性。新兴的人工智能技术如何影响未来国际政治、权力格局和全球战略稳定性？人工智能的军事化发展如何影响大国兴衰的历程？世界是否会出现新的人工智能军事霸主？致命性自主武器系统的发展，是否会如同开启了潘多拉魔盒一般，将人类社会拖入恐怖的终结时代？人类是否能够依靠理性的力量构建人工智能的军控机制，有效规避自主武器扩散带来的安全风险？一切都是未知之数。

人们已经意识到，人工智能领域的军备竞赛是不可避免的，基于道德、法律和战略安全等方面的考量，保持人类对人工智能军事化发展进程的控制权，就成为一种理性和现实的选择。盖斯特认为："由于人工智能研究与军事应用之间的长期联系，放弃

人工智能军备竞赛为时已晚。我们需要在管理有序、安全可靠的人工智能军备竞赛与管理不善、可能导致灾难性后果的军备竞赛之间做出选择。”① 当前，围绕人工智能如何实现良好、安全的军事化发展这个问题，存在多种治理方案。例如，联合国《特定常规武器公约》会议致命性自主武器政府专家组提出，所有武器都应受到“有意义的人类控制”。② 沙瑞尔提出“创建保持人类判断在战争中作用的通用准则”。③ 美国国防创新委员会发布的《人工智能原则：国防部人工智能应用伦理的若干建议》，就美军如何合乎伦理和法律地使用人工智能技术，提出“责任、公平、可追踪、可靠、可控”五大原则。④

上述方案在允许研发、部署人工智能武器的前提下，通过技术、政策等手段为人工智能武器确立行为准则，来限制人道主义灾难的发生，其作为军备控制原则具有一定的宏观指导意义，但在实践层面，如何实现“有意义的人类控制”？综合考虑人工智能系统和战争系统的双重复杂性和不可控性，在战争实践中，人工智能武器能否准确理解、严格执行人类指令本身就具有不确定性，人类也往往难以清晰地认识高度复杂的人工智能“技术黑

① GEIST E M. It's already too late to stop the AI arms race—We must manage it instead [J]. Bulletin of the Atomic Scientists, 2016, 72 (5): 318.

② RIEBE T, SCHMID S, REUTER C. Meaningful human control of lethal autonomous weapon systems: the CCW - Debate and its implications for VSD [J]. IEEE Technology and Society Magazine, 2020, 39 (4): 36.

③ 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：400.

④ Defense Innovation Board. AI Principles: Recommendations on the ethical use of artificial intelligence by the Department of Defense: supporting document [Z]. United States Department of Defense, 2019.

箱”。因此，构建“人在回路”的军事系统结构，并不必然地实现对人工智能系统的“有意义的人类控制”。

驾驭人工智能时代的安全挑战需要更为系统的伦理探索，并提出更为具体、更具有可行性的伦理道德规范。本书尝试提出对于新兴人工智能战争的伦理评估框架。

首先，梳理了人工智能军事化的发展历程和发展态势，揭示了人工智能军事应用带来的安全风险，在此基础上从情报搜集、分析决策和火力打击三个方面，系统分析了智能化战争整个作战流程对正义战争伦理原则的挑战。（1）在情报搜集方面，智能化战争发起国将网络空间作为情报搜集的新战场，其实质是以网络手段挑战其他主权国家在网络空间治理方面的合法权威，进而导致一系列伦理危机。一方面，网络情报战通过互联网和移动存储介质传播病毒，它的感染范围与破坏程度并不完全取决于攻击者的意图与作战规划，可能会造成大量民用计算机和网络系统感染，难以符合区分原则和相称性原则的要求。另一方面，由于无法准确追溯网络攻击源头，网络情报战在战争责任分配和承担重建义务等方面也面临困境。（2）在分析决策方面，大数据技术被用来对情报信息进行分析、加工处理，研判袭击行动规律，并为随之展开行动提供决策参考。大数据分析手段在智能化军事行动中的广泛应用，虽然在一定程度上提升了锁定目标和预判袭击事件的效率，但也导致军方过度依赖大数据手段和数据量化标准，引发先入为主的错误区分和大数据反恐的滥用。（3）在火力打击方面，智能化作战平台对己方的“零伤亡”与对敌方的“定点清除”式精确打击，在一定程度上契合了正义战争理论的道德准

则，但是，对于实施无人机反恐作战的国家而言，无人化反恐的低成本、零风险，使其能够毫无顾忌地跨越战争门槛，在战争与和平的抉择中不断挑战更低的道德界限。

其次，从正义战争理论的分析框架出发，以美军无人机刺杀伊朗“圣城旅”指挥官苏莱曼尼的作战行动为例，深入揭示刺杀行动在遵循开战正义、交战正义原则方面存在的问题，具体观点体现为：（1）从美军刺杀苏莱曼尼的理由来看，无论是针对恐怖主义袭击行使自卫权，还是针对潜在恐怖主义威胁行使预先防卫权，都是欺骗、愚弄国际社会和国内民众的拙劣借口。（2）从刺杀过程来看，美军使用无人机刺杀苏莱曼尼的过程，远非特朗普政府宣称的那样在道德上“完美无瑕”。为了达成在第一时间击毙苏莱曼尼的目的，特朗普政府丝毫没有考虑其他同车人员的身份识别和攻击合法性问题，以及刺杀行动对中东地区局势和民众生命财产安全的一系列负面影响。（3）从根源上分析，新保守主义的道德观念是主导美国“反恐战争”逻辑的道义基础。虽然特朗普政府粉饰其军事行动的正义性，但是无论其给出的理由如何正当、动机如何高尚，本质上都体现了在新保守主义思想驱动下美国政府的政治与战略考量。

再次，揭示了智能化战争时代“慈化”愿景的实践困境，具体体现为：（1）作战环境存在现实与虚拟的两难困境。智能化战争中的虚拟战场环境，缩短了作战人员与作战目标之间的物理视距，同时也拉大了两者之间的精神距离。在虚拟环境下，武器的操作人员日益与他们的行为所产生的后果分割开来，其结果是抑制了作战人员对杀戮行为的愧疚感，进而威胁到他们作为军人的

专业道德和作为人类的基本良知。(2) 作战手段的运用存在滥杀与精杀的两难困境。这类武器精确的“定点清除”，使得反恐作战人员在面对平民和民用目标时不再谨慎小心，出于这一心态作祟，美军不断增加无人机在反恐作战中的使用频度，将其滥用于军民目标混杂的反恐军事行动之中，从而造成对区分原则的进一步弱化。(3) 作战结果预测存在未知与已知的两难困境。由于作战结果由未知变为已知，智能化打击方式在西方政治家眼中已经成为一件几乎无风险的机器，战争的胜利也就成为一个可以自我实现的预言。依靠科技优势来实现战争的人道主义愿景，已经背离了人道主义的道德主题，成为西方政治家缓解舆论压力和实施战争动员的工具。

最后，从技术、哲学、法律等方面，就规制智能化战争中违背人道主义的行为提出了一系列对策建议：(1) 就技术规制进路而言，主要体现在人工智能控制系统的技术设计过程中，依靠计算机编程的方式，使之根据一组固定的道德规则运行，从而符合战争伦理和战争法的规定，即将自上而下的道德预置与自下而上的道德生成相互结合的技术进路。在利用编程预先设置基本规则的基础上，通过真实战场环境“改进—继续—改进”的机器学习过程，不断修正既有算法偏见，实现区分目标和相称性评估能力的持续增长，使无人作战平台在战场的行动符合人道主义精神和战争规约要求。(2) 就哲学规制进路而言，军用人工智能系统的设计研发，不仅要考虑交战规则的预置，还必须充分考虑隐含其中的伦理价值观念，将正义战争所代表的人道主义价值观，预先嵌入基础算法和程序的设计中，其本质是坚持将智能化战争的控

制权牢牢掌握在人类手中，确保人类不受制于机器。(3) 就法律规制进路而言，应通过明确的法律规定限制智能化战争行为。首先，自主武器系统使用的目的应限于防卫而非进攻，即限定于执行纯粹防卫性质的军事行动，诸如对恐怖主义活动的侦察监视、情报搜集、危机预警和应急救援等非攻击性军事任务，而不宜将其使用在大规模战争或易于造成误判、误伤的“定点清除”或“斩首”行动之中。其次，自主武器系统使用范围限于非平民聚居的交战区域，禁止在人类活动频繁的区域特别是人口密集的生活区使用无人机进行反恐作战。最后，自主武器系统的使用程度限于非致命，尽可能对恐怖分子嫌疑人使用非致命武器。(4) 就军控规制进路而言，必须依托联合国这一全球性国际安全机构，全面统筹协调各方诉求，切实加强国际合作，适时、恰当地推进构建人工智能军控体系，通过友好协商的方式，就强化人工智能的军备控制达成最广泛的战略共识，推动相关全球治理体系的搭建。为此，各国必须把握当前的战略机遇期，在联合国框架下共同推动军控的各项工作，共同敦促军事强国控制军用人工智能技术的发展。(5) 就舆论规制进路而言，首先是在科学共同体内部营造遵循科研伦理和履行研发者主体责任的舆论氛围，其次是发动整个国际社会形成对使用致命性自主武器的舆论攻势，最后是完善军用人工智能技术的全球舆论规制行动者网络。

总体而言，在学术价值方面，本书进一步深化了对高技术战争条件下战争伦理评判原则和依据的探讨，聚焦网络攻防技术、大数据技术及人工智能技术军事应用对现代战争伦理体系的冲击，对“智能化战争在开战正义层面是否具有正当理由、正当意图和

合法权威”“智能化战争在交战正义层面是否遵循区分原则和相称性原则”“智能化战争在战后正义层面是否妥善实现战争责任追究”等问题进行了伦理评判。在应用价值方面，本书就如何规制智能化战争中一系列违背人道主义精神的问题，从技术、哲学、法律、军控和舆论五个方面提出了具体的规制进路，具有一定的实践指导意义。

作 者

2022 年 8 月

目 录

第一章　人工智能军事化发展历程与态势分析 / 1

一、从无人机到自主武器系统 / 3
二、美国军用人工智能发展态势与动因 / 14
三、其他国家和组织加速推进军用人工智能研发和应用 / 30

第二章　人工智能军事化的国际安全影响 / 34

一、激化战争失控的风险 / 35
二、加剧军备竞赛的安全困境 / 41
三、威胁全球战略稳定性 / 49
四、挑战战争伦理规约 / 62
五、应对安全风险的启示 / 84

第三章　信息搜集与网络情报战 / 91

一、信息搜集和网络情报战的途径方式 / 93
二、网络情报战与开战正义 / 96
三、网络情报战与交战正义 / 100
四、网络情报战与战后正义 / 105

第四章　分析决策的正义性：以大数据分析为例 / 109

一、大数据分析的军事价值及应用 / 110
二、大数据分析的伦理挑战 / 114
三、大数据分析如何合乎道义？ / 118

第五章　无人机“定点清除”行动 / 121

一、美国无人机“定点清除”行动的发展 / 122
二、无人机“定点清除”行动正义性问题的研究综述 / 127
三、智能识别方式下火力打击的区分困境 / 142
四、“平台无人”模式下的武力控制困境 / 149
五、“机器 VS 人”攻击模式下的责任困境 / 154
六、无人作战的人机融合：挑战与出路 / 157

第六章　案例分析：以美军刺杀苏莱曼尼为例 / 171

一、美国定义刺杀事件的“正当理由” / 173
二、刺杀事件是否遵循开战正义的伦理原则？ / 177
三、刺杀事件是否遵循交战正义的伦理原则？ / 183
四、美国“反恐战争”的非正义性分析 / 188

第七章 智能化战争“慈化”愿景及其实践困境 / 195

一、智能化战争“慈化”愿景 / 196
二、作战环境存在现实与虚拟的两难困境 / 198
三、作战手段运用存在滥杀与精杀的两难困境 / 202
四、作战结果预测存在未知与已知的两难困境 / 205

第八章 智能化战争风险的规制进路 / 211

一、技术规制进路 / 213
二、哲学规制进路 / 215
三、法律规制进路 / 218
四、国际军控规制进路 / 221
五、舆论规制进路 / 226

结 语 / 229

参考文献 / 236

附 录 / 242

后 记 / 244

第一章
人工智能军事化发展历程与态势分析

自2014年11月美国国防部提出“第三次抵消战略”以来[①]，国际社会以人工智能为主题的安全事件不断涌现。2015年12月，在叙利亚拉塔基亚754.5高地作战中，俄罗斯军队首次将整建制的机器人部队投入实战[②]。2020年1月2日，美军使用“捕食者”无人机刺杀伊朗“圣城旅”指挥官苏莱曼尼，开以远程遥控方式击毙主权国家军政领导人之先河[③]。2020年3月，在利比亚政府军与反对派武装的战斗中，一架土耳其卡古-2四轴飞行器使用自主攻击模式

① LOUTH J，TAYLOR T. The US Third Offset Strategy [J]. The RUSI Journal，2016，161(3)：66.

② 张晏. 无人化：审视智能化战争窗口 [N]. 解放军报，2018-11-29 (11).

③ ANKARA. U. S. killing of Iranian commander threatens regional stability：Turkish FM[EB/OL]. (2020-01-07)[2022-04-07]. http://www.china.org.cn/world/Off_the_Wire/2020-01/07/content_75584623.htm.

猎杀人类目标，这是战争史上第一次在没有人类指令情况下发生的无人机自主攻击事件。[①] 上述里程碑事件反映了一个共同的主题：人工智能的军事化趋势不可逆转，智能化战争的时代已经来临。

人工智能的军事应用在更广的范围、更深的层次上不断拓展，由此产生的安全影响引发国际社会的高度关注和巨大争议。以美国佐治亚理工学院的阿金为代表的技术乐观主义者认为，人工智能可以替代人类参加战争，确保军人“零伤亡”。同时，人工智能还具有人类无可比拟的战场观察能力，可以提高打击精度，降低对平民的附带伤害。[②] 与之针锋相对，以霍金、马斯克为代表的顶尖科技专家对人工智能的军事化发展表达了忧虑和恐惧。2018 年，在瑞典斯德哥尔摩召开的国际人工智能联合会议上，包括特斯拉公司首席执行官马斯克、谷歌 DeepMind 公司三位联合创始人在内的 2000 多名人工智能领域的学者共同签署了《致命性自主武器宣言》，倡议禁止研发具有致命能力的自主机器人系统。[③]

人工智能的军事化发展究竟会以怎样的方式影响我们身处的安全环境和未来的安全前景？它会带来更人道的作战方式和不流血的战争，还是如同开启潘多拉魔盒一般，将人类社会彻底拖入恐怖的终结时代？人工智能军事应用前景，成为国际安全研究领域的一个焦点议题。目前，关于人工智能的国际安全影响，国内学界除傅

① MORAN R. Meet the future: drone hunts down a human without being ordered to[EB/OL]. (2021 - 05 - 31) [2022 - 04 - 07]. https://www.sgtreport.com/2021/05/meet - the - future - drone - hunts - down - a - human - without - being - ordered - to/.

② ARKIN R C. The case for ethical autonomy in unmanned systems [J]. Journal of Military Ethics, 2010, 9 (4): 333.

③ Future of Life Institute. Lethal autonomous weapons pledge[EB/OL]. (2018 - 06 - 06) [2022 - 04 - 07]. https://futureoflife.org/lethal - autonomous - weapons - pledge/.

莹[①]、封帅[②]等少数学者从人工智能与安全的宏观视角切入这一议题外，聚焦军用人工智能的专题性研究和深度剖析仍暂付阙如。国外学界就这一专题取得的研究成果更为丰富。然而，囿于国情、军情、价值观念以及研究角度的差异性，加之当前人工智能技术及其在军事应用领域的持续快速发展，现有研究在系统性、深入性和针对性方面仍存在较大局限。本章梳理人工智能军事化的发展历程，并对全球军用人工智能发展态势进行分析研判。

一、从无人机到自主武器系统

从军事技术发展演化的动态视角来看，人工智能的军事化发展是一个逐渐走向自主化的过程。

（一）无人机

作为军事智能化的先声，无人机发展的历史与载人飞机发展的历史一样长。早在20世纪初期，欧美国家已经开始探索利用无线电遥控方式操控无人飞行器。20世纪上半叶，无人机主要作为飞行员训练的靶机使用，其目的是提升飞行员的空中作战技能。在20世纪60年代的越南战场，美军人力侦察应对越南军队在热带雨林地区的游击作战中付出了巨大的伤亡代价。因此，美军开始尝试将无人侦察机大规模应用于军事领域，替代传统侦察部队完成危险环境下的侦察任务。总体而言，在整个20世纪，无人机的军事应用进程相对缓慢，其对战争胜负的影响也相对有限，尚未成为一种具有独立改

① 傅莹. 人工智能对国际关系的影响初析［J］. 国际政治科学，2019，4（1）：1－18.

② 封帅. 人工智能时代的国际关系：走向变革且不平等的世界［J］. 外交评论，2018，35（1）：128－156.

变战局能力的颠覆性武器装备。

进入21世纪，伴随着人工智能技术的军事应用，一批无人化作战平台和作战装备出现并不断发展。在当今世界各国竞相研发无人机的浪潮中，美国无人机研发和军事应用的水平居于全球领先地位，特别是“捕食者”和“收割者”无人机，成为美军执行“定点清除”行动的重要工具。

无人机之所以能够成为美军应用于“定点清除”行动的利器，主要是基于两个方面的原因。一是“9·11”事件以后美军作战的现实需求。事实上，早在20世纪90年代中期，美国中情局就开始操纵无人机监视和侦察敌对分子的行踪，并曾经多次搜索到本·拉登等恐怖分子的行踪，然而，由于缺乏及时有效的打击手段，美军无法在第一时间执行“斩首”任务。2001年，“9·11”事件发生以后，美国以阿富汗战争为起点，在全球范围内发动了声势浩大的“反恐战争”。在打击分散在世界各地的恐怖分子的过程中，美军作战部门面临的一个关键问题是，如何在一个高度机动、混杂、变幻的战场环境中，迅速、准确地发现恐怖分子并下达攻击指令，保证对恐怖分子的“发现即摧毁”。上述问题对综合化“察打一体”的空中无人反恐作战平台提出了紧迫的需求。以“捕食者”和“收割者”为代表的“察打一体”式无人机的问世，彻底改变了无人作战的“观察—判断—决策—行动”（OODA）循环，使人类战争进入“发现即摧毁”的“秒杀时代”。二是美军智能化装备技术体系的发展与完善。早在科索沃战争时期，美军就已经装备了“捕食者”无人机的原型机，但是这种无人机只有战场侦察能力，不具备火力打击能力。“9·11”事件以后，为配合阿富汗作战需要，美军凭借领先世界的军事技术整体优势，发展出了集情报侦察与火力打击于一体的无人化远程遥控作战平台。曾任美国空军参谋长的江珀将军回

忆了“捕食者”无人机的研发历程：“捕食者”无人机在设计之初没有预想到会被武装，它的有效载荷不到500磅，仅能装备重量最轻的标准“哑弹”，而激光制导导弹GBU－24是“哑弹”重量的4倍。解决方案是装备激光制导的AGM－114“地狱火”，这是一种用在陆军和海军陆战队攻击直升机上的空对地反坦克导弹，每枚重量约100磅，是对付“不需要大弹头就能除掉的那些稍纵即逝的目标”的理想手段。2001年2月16日，“捕食者”在内利斯空军基地的测试过程中成功发射了一枚AGM－114C“地狱火”导弹并击中目标。让“捕食者”变成真正捕食者的愿望终于实现了。①

在此后美军的反恐作战中，无人机迅速发展成为具有侦察打击一体化功能的反恐利器，在美军遍布全球的“定点清除”行动中展现出越来越重要的作用。2001年11月，装备“地狱火”导弹的“捕食者”无人机在美军击毙本·拉登心腹阿提夫的行动中扮演重要角色；2002年11月4日，在阿富汗战场之外的也门，美军使用“捕食者”无人机向恐怖组织目标发射精确制导炸弹，击毙“基地”组织重要头目阿尔哈特希。美国军事历史学家耶讷还原了刺杀行动的场景：在匪徒们所乘坐的SUV上空10000英尺处，一架身形瘦长、灰白色的飞机正在以比他们稍快的速度飞行。通过这架无人飞行器的视频链路，在约350千米外的吉布提，一个坐在某个控制中心的美国人已经观察这辆SUV一个小时了。突然一道火焰从这架缓缓飞行的飞机机翼上喷射出来，一条云状尾迹快速飞向沙漠地面。六个匪徒感到一阵可怕的爆炸震动这辆SUV，他们立即被一团猛烈燃烧

① 耶讷. 无人机攻击［M］. 刘春辉，赵希源，译. 北京：中国市场出版社，2014：128.

的火球吞没。[①] 这是首次由一架武装“捕食者”无人机在阿富汗战区以外袭击作战目标的行动。

以击毙“基地”组织重要头目阿尔哈特希为契机，各国政府和军方开始重新审视和评估武装无人机在“定点清除”行动中所扮演的角色，并确认无人机反恐拥有传统作战模式难以比拟的优势。相对于有人作战飞机，无人机造价相对低廉，具有较强的隐秘性和方便高效的操作性能，便于回收，安全风险小，不易威胁到士兵的生命安全，同时具有极强的环境适应能力，可在各种复杂地形和气象环境下连续飞行，实现超长待机。因此，无人机日益成为“定点清除”行动的最优选择。此外，在常规战争模式下，后勤问题始终是影响和制约军事行动展开的重要因素。“察打一体”无人机以其独有的“搜索－猎杀”模式，颠覆性地突破了常规空战需要地面部队配合的作战模式，将人从现实战场环境中彻底解放出来。无人机操控员能够在千里之外发起远距离打击，精确实施“外科手术刀”式“斩首”攻击，不仅可以实现己方军事力量的“零伤亡”，而且可以一劳永逸地解决后勤补给受作战距离和时间限制的问题。此外，无人作战方式也极大地缩减了战争动员与准备环节所必需的成本。“察打一体”无人机配备的“地狱火”导弹具有快速精确制导的能力，使“定点清除”行动完全能够在对方做出反应之前即告完成。基于以上技术优势，无人机“定点清除”行动日益成为军事领导人偏爱的军事打击方式。

在利比亚内战中，利用无人驾驶飞行器实施“定点清除”攻击的优势逐渐显现出来。加迪指出：利比亚“国民军”和利比亚民族

① 耶纳. 无人机攻击［M］. 刘春辉，赵希源，译. 北京：中国市场出版社，2014：Ⅱ－Ⅲ.

团结政府作为交战双方，都在军事行动中接受无人机“定点清除”的外部支持，使用无人机在利比亚首都的黎波里郊区的战斗中进行情报搜集、监视和侦察，以及对关键目标进行“定点清除”攻击。依据联合国统计数据，在6个月中，双方共执行了约900架次无人机作战任务，平均每天约5架次。其中，约600架次由利比亚“国民军”执行，约300架次由利比亚民族团结政府执行。双方利用无人机精确攻击了对方的无人机操作站点和基础设施。利比亚民族团结政府声称其在此期间所遭受的伤亡中有多达三分之二是由无人机打击造成的。①

在2020年爆发的纳卡冲突中，阿塞拜疆使用了大量无人机通过“定点清除”攻击模式攻击亚美尼亚的军事目标，作为双方争夺焦点的纳卡地区俨然成了新一轮无人机大战的战争试验场。据环球网报道，在这次冲突中，阿塞拜疆军方主要通过使用两种无人机来实施“定点清除”攻击。一种是依靠土耳其提供的TB2“察打一体”无人机，这种无人机飞行高度为8000多米，作战高度为5000多米，时速为100多千米，能在翼下挂载4枚半主动激光制导空地导弹，最大射程4千米，能在战场上空长时间盘旋，择机发起致命攻击。阿塞拜疆使用该无人机实施“定点清除”攻击，精准摧毁了亚美尼亚的坦克装备车、地面防空系统、火炮、仓库等。另一种无人机是由以色列兵工企业制造生产的“哈洛普”自杀式无人机。阿塞拜疆军方通过遥控方式使用“哈洛普”对地面目标实施“定点清除”攻击，具有极高的精确性。在交战过程中，“哈洛普”主要用于精准攻击亚美尼亚的俄式S300导弹系统。

① GADY F-S. Useful, but not decisive: UAVs in Libya's civil war [EB/OL]. (2019-11-22) [2022-04-07]. http://www.iiss.org/online-analysis/online-analysis/2019/11/mide-uavs-in-libyas-civilwar/.

（二）自主武器系统

客观而言，远程遥控无人机“定点清除”的作战方式是介于信息化作战方式和智能化作战方式之间的过渡阶段。当前，世界主要军事强国智能化军备竞赛愈演愈烈；未来，完全智能化的自主武器系统将有可能成为作战的重要打击平台。2015 年 7 月 28 日，国际人工智能联合会议发表全球 116 名科技专家的联名公开信，物理学家霍金、特斯拉公司首席执行官马斯克、苹果公司联合创始人沃兹尼亚克等人共同指出：致命性自主武器（“杀手机器人”）有可能成为军事史上的第三次革命。一旦它们发展起来，可能导致比以往任何时候更大规模的武装冲突，而且速度比人类所能理解的还要快。[①] 自主武器系统大规模应用于军事领域，将取代人类的战争主体地位，掌握生杀予夺的至高权力，进而对现有的战争伦理体系构成新的挑战。

当前，学界普遍认为，自主性是定义自主武器系统的典型标签。自主武器系统的定义，政府部门和研究者主要从武器系统的自主性方面进行界定。如美国国防部出台的发展自主武器的官方文件《国防部指令 3000.09：武器系统中的自主性》，指出自主武器系统是一种“在激活后无须进一步人工输入就可以自主选择目标和交战的武器系统”。虽然美国国防部对自主武器系统给出了官方界定，但是当前学界对于自主性在武器系统中的具体体现，仍然存在一定的分歧与争议。格里夫等人认为，自主武器系统是一种“算法决定一切、

① Future of Life Institute. Autonomous weapons: an open letter from AI & Robotics Researchers [EB/OL]. (2016 - 02 - 09) [2022 - 04 - 07]. https://futureoflife.org/open - letter - autonomous - weapons/.

自主行动且以无须人类干预的方式运行的武器系统”。[①] 霍罗威茨进一步提出，可以从自我机动性、自我指导性和自我决定性三个方面来衡量武器系统的自主性程度。[②] 红十字国际委员会将自主武器系统定义为“可以独立选择和攻击目标的武器，即在获取、跟踪、选择和攻击目标等关键功能方面具有自主权”。[③] 沙瑞尔在《无人军队：自主武器与未来战争》一书中将自主武器系统定义为“能够自主独立完成全部作战任务周期，即搜索目标、决定攻击和发动攻击的武器”。[④]

综合以上定义，笔者认为，真正意义上的自主武器系统是一种拥有完全自主权的杀人机器。在人工智能技术的驱动下，它能够自主完成整个致命攻击的流程，即自主运用传感器搜索潜在目标的数据信息，通过数据分析自主识别、锁定目标，自主评估使用武力的后果，自主做出攻击决策。

依据当前美国等世界军事强国在提升武器装备自主性方面的具体实践，当前自主武器系统自主性能的发展集中在以下三个方面。

一是自主情报分析。情报分析的失误可能导致危机的升级甚至战争的爆发。传统的情报分析主要基于人力系统来完成。自主武器系统可以对海量情报数据信息进行人工智能分析。为实现自动化情报处理，2017 年，美国国防部宣布成立了代号为 Maven 的算法战跨

① GREEF T D , DONGEN K V , GROOTJEN M , et al. Augmenting cognition: reviewing the symbiotic relation between man and machine [J]. Foundations of Augmented Cognition, 2007, 4565: 439 - 448.

② HOROWITZ M C. When speed kills: lethal autonomous weapon systems, deterrence and stability [J]. Journal of Strategic Studies, 2019, 42 (6): 764 - 788.

③ SHARKEY A. Autonomous weapons systems, killer robots and human dignity [J]. Ethics & Information Technology, 2019, 21 (2): 75.

④ 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：57 - 58.

职能小组，研究在情报搜集单元实现计算机视觉和机器学习算法的融合，其目的是“将大量动态形式的视频数据转化为可用、可操作的情报和见解”。[①] 例如，在针对“伊斯兰国”的反恐作战中，自主武器系统利用算法筛选无人机录像画面的高价值信息，从而自动识别敌对目标的活动。自主武器系统可以将当前人类分析员需要花费数小时筛选录像画面以获取有价值信息的工作转变成几分钟的自动化处理工作。

二是自主作战指挥与组织编成。利用人工智能算法，自主武器系统可以基于对战场的实时分析，依据战场态势发展进行动态任务规划，自主生成可行的作战行动路线方案，从而提高作战响应的时敏性和战时指挥决策的质量与速度。2017 年，美国国防高级研究计划局提出的“马赛克战争”就是一个典型的例子。它将自主武器系统、低成本传感器网络、联合多域命令和智能控制节点组合成为一个具有高度适应性的作战网络，依据战场情况和作战需求的变化发展，动态生成作战指挥命令，快速实现作战形态的自我聚合与分解，为作战指挥提供前所未有的自治性和灵活性。[②] 随着作战决策的加速以及对决策的加速再响应，人类分析和理解战场态势的速度将跟不上自主武器系统运行的速度。

三是自主火力攻击。自主武器系统能够在不依赖外界指令和设备支持的情况下，在复杂作战环境中依靠人工智能系统独立做出开火决定。虽然对于目前大部分人类开发的半自主武器系统而言，远

① SUCHMAN L. Algorithmic warfare and the reinvention of accuracy [J]. Critical Studies on Security, 2020, 8 (2): 175 - 176.

② HAYSTEAD J. DARPA's mosaic warfare moving to address the ever - more - rapidly - paced advances/changes in fielded threat capabilities [J]. Journal of Electronic Defense, 2020 (2): 43.

程控制中心的操控员仍然掌握开火的最后按钮，但是一些军事强国已经研制出具备自主开火模式的完全自主武器。如英国军方研发的“硫磺石”导弹就具备“发射后不管”的极高自主性，操作人员只需为“硫磺石”导弹设置一个大致的杀伤范围，它就会向该区域内的敌方目标发起自主攻击，即每枚导弹可自行决定攻击哪辆坦克或其他作战车辆。[①] 韩国军方部署在朝韩边境的 SGR－A1“哨兵”机器人，通过使用模式识别软件和弱光摄像头识别、锁定越境者，并在必要的时候自主掌握攻击权限。[②]

需要说明的是，自主武器系统的发展存在一个自主性程度逐渐提高的过程，即从远程遥控的半自动系统向完全自治系统发展的演进过程。自主武器系统是人工智能技术推动武器装备创新发展的典型案例之一，它在军事领域的横空出世，引发了“人－武器”相互关系和战斗力生成机理的颠覆性转变，其发展历程体现了武器系统“非自主—半自主—完全自主”的演进趋势。

在阿富汗战争中，美国使用“捕食者”无人机对“基地”组织和塔利班头目进行“定点清除”式攻击，揭开了无人化“反恐战争”的序幕。在此后的 10 余年间，“基地”组织重要头目阿尔哈特希、“自由者大会”头目霍拉萨尼、巴黎《查理周刊》恐怖袭击事件策划人安西、“伊斯兰国”头目巴格达迪、伊朗“圣城旅”指挥官苏莱曼尼等人皆命丧无人机之手。作为美军执行“定点清除”行动的利器，“察打一体”的“捕食者”和“收割者”无人机，实质上是一种有人监督的半自主武器系统，已经具有自主导航定位、自

① 沙瑞尔. 无人军队：自主武器与未来战争［M］. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：116－117.

② KUMAGAI J. A robotic sentry for Korea's demilitarized zone［J］. IEEE Spectrum, 2007, 44（3）：16－17.

主目标识别和半自主火力打击的能力，人类实时监控无人作战系统的运行，可以在必要时候对武器系统进行干预和接管。

当前，智能化作战系统的发展正处于由“平台无人、体系有人”的半自主武器向完全自主武器加速发展的阶段。为谋求新一轮军事变革的可持续竞争优势，各国纷纷加速推进人工智能技术的军事应用进程。武器装备的自主化程度不断提升，催生了以完全自主武器系统为代表的一大批致命性自主武器系统。完全自主武器系统的出现，彻底颠覆了人类作为武器操纵者和致命行为执行者的历史。2017 年，国际非政府组织“杀手机器人禁令运动”对外公布了一段视频。一群由人工智能技术驱动的机器“杀人蜂”闯入教室，通过人脸识别方式，将十余名正在上课的学生瞬间击杀，场面极度血腥。虽然视频中的杀人场景是由计算机模拟而成，但在现实世界中，各军事强国围绕自主武器系统的军备竞赛已在如火如荼地展开。早在 2012 年 11 月，美国国防部就出台了发展自主武器的官方文件《国防部指令 3000.09：武器系统中的自主性》，规定了研发半自主和自主武器的通用原则，并实施了联合无人空战系统、高自主性远程反舰导弹、X－47B 无人机等一系列自主武器技术和装备研发项目。[①] 与美国自主武器的发展相同步，在叙利亚战争中，俄罗斯军方就开始将整建制的半自主机器人投入战场。除美、俄以外，英国的“雷神”无人机和“硫磺石”导弹、法国的“神经元”无人机、印度的“奥拉”无人机以及韩国的 SGR－A1“哨兵”机器人等，都是具有自主作战能力的武器系统。

① 沙瑞尔. 无人军队：自主武器与未来战争［M］. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：111－130.

当前各国自主武器系统研发、应用情况一览表①

国家	名称	自主程度	研制、试验、作战应用情况
美国	MQ－1“捕食者”无人机	半自主	自2001年以来，广泛应用于美军“定点清除”式反恐作战中
美国	MQ－9“收割者”无人机	半自主	“捕食者”无人机升级版，参与对伊朗“圣城旅”指挥官苏莱曼尼等人的“定点清除”攻击
俄罗斯	“平台－M”陆战机器人	可选的半自主/完全自主模式	2014年随俄军参加了城市作战演习，2015年在叙利亚战场配合叙利亚政府军作战
俄罗斯	“阿尔戈”战斗机器人	可选的半自主/完全自主模式	2015年，在叙利亚战场参与了围攻拉塔基亚754.5高地的作战行动
英国	“硫磺石”反坦克导弹	可选的半自主/完全自主模式	2013年，处于毫米波自主模式的“硫磺石”导弹完成了世界上第一次齐射打击试验
英国	“雷神”无人机	可选的半自主/完全自主模式	由BAE系统公司研制，设置了有主动攻击目标的可选模式，可在战场上自主搜寻和攻击敌对目标

① 沙瑞尔. 无人军队：自主武器与未来战争［M］. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：111－130.

续表

国家	名称	自主程度	研制、试验、作战应用情况
印度	“奥拉”无人机	可选的半自主/完全自主模式	“奥拉”战术隐身无人机项目于2009年启动，2023年完成飞行样机的研制
韩国	SGR－A1“哨兵”机器人	可选的半自主/完全自主模式	由韩国三星公司研制，目前部署在朝韩边境执行警戒和打击越境者等任务

二、美国军用人工智能发展态势与动因

伴随以深度学习、神经网络为标志的人工智能技术进入产业化发展阶段，世界军事强国日益充分地意识到人工智能对于国防和军队建设的引领作用。2014年，美国国防部提出了以人工智能为核心、以改变未来战争形态的颠覆性技术群为重点的“第三次抵消战略”，宣布将“机器人技术、自主操作和控制系统、先进的计算和大数据”等人工智能技术作为“第三次抵消战略”的基石[①]。“第三次抵消战略”的出台成为全球军用人工智能发展的标志性事件，不仅直接推动了美军面向未来战争的新军事变革进程，而且引发了其他传统军事强国和新兴国家的跟随与效仿。

美国军用人工智能的发展始于奥巴马执政时期。特朗普执政时

① WORK B. The Third U. S. Offset Strategy and its implications for partners and allies [EB/OL]. (2015－01－28) [2022－04－07]. https://www.defense.gov/Newsroom/Speeches/Speech/Article/606641/the－third－us－offset－strategy－and－its－implications－for－partners－and－allies/.

期，在“美国优先”这一核心执政理念的驱动下，军用人工智能发展得以持续推进。拜登当选美国总统后，承诺对提升美国竞争力的关键技术加大投资力度，更为重视且不遗余力地推进军用人工智能发展进程，不仅将人工智能技术视为维护国家安全和打赢未来战争的关键技术之一，而且刻意强调其他竞争对手在人工智能领域对美国构成的军事威胁，力图通过谋取人工智能领域的领先地位来遏制、打压竞争对手，巩固全球霸权。

（一）发展态势

1. 构建军用人工智能发展战略体系

早在奥巴马执政时期，美国政府和军方注重从战略层面对军事智能化进程进行设计筹划。从2016年10月到2016年12月，短短三个月的时间，聚焦人工智能领域，奥巴马政府密集颁布了《准备迎接人工智能的未来》[①]《国家人工智能研发战略规划》[②]和《人工智能、自动化与经济》[③]三份战略发展报告，从战略高度谋划国家人工智能领域的中长期发展，并将发挥人工智能在维护国家安全与国防建设中的作用列为重要内容。其中，《国家人工智能研发战略规划》从战略选择的角度，明确了美国人工智能研究与开发的重点方向，将人工智能对于“强化国家安全和国土安全”的作用，放在与

① Executive Office of the President. Preparing for the future of Artificial Intelligence [EB/OL]. (2016-10-12) [2022-04-07]. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

② National Science and Technology Council. The National Artificial Intelligence Research and Development Strategic Plan [EB/OL]. (2016-10-13) [2022-04-07]. https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf?ref=innovationpolicy.org.

③ LEE K. Artificial Intelligence, Automation, and the Economy [EB/OL]. (2016-12-20) [2022-04-07]. https://obamawhitehouse.archives.gov/blog/2016/12/20/artificial-intelligence-automation-and-economy.

“促进经济繁荣”和“改善教育机会和生活质量”同等重要的地位，并作为促进美国发展的优先事项来对待。[①]《准备迎接人工智能的未来》就人工智能军事应用的潜在领域进行了较全面的战略评估，指出：“人工智能有可能在一系列与国防有关的活动中扮演重要角色。后勤、维修、基地行动、退伍军人医疗保健、战场医疗援助和伤员后送、人员管理、导航、通信、网络防御和情报分析等非致命性活动可受益于人工智能，从而使美军行动更安全、更有效率。”[②] 此外，报告专辟“武器系统中的人工智能”“网络安全中的人工智能”等章节，就如何实现人工智能在维护国家安全中的潜在价值提出咨询建议。

较之奥巴马执政时期，特朗普政府在“美国优先”这一核心执政理念的驱动下推进军事智能化进程。2017 年 12 月 18 日，特朗普总统签署的《国家安全战略》指出：“美国面临的安全环境受到技术进步和不断变化的安全特征的影响。面向战争开发新技术的动力是现实的，且以更低的进入门槛向更多的参与者扩展，并以更快的速度前进。新技术包括高级计算、大数据分析、人工智能、自主性、机器人、定向能、高超音速和生物技术，这些正是确保美国能够打赢未来战争的技术。”[③]

奥巴马执政时期，虽然政府已经从战略高度重视军事智能化的发展，并且颁布了一系列官方文件，但就性质而言，基本停留在战

① National Science and Technology Council. The National Artificial Intelligence Research and Development Strategic Plan [EB/OL]. (2016 - 10 - 13) [2022 - 04 - 07]. https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf?ref=innovationpolicy.org.

② Executive Office of the President. Preparing for the future of Artificial Intelligence [EB/OL]. (2016 - 10 - 12) [2022 - 04 - 07]. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

③ Department of Defense. Summary of the 2018 National Defense Strategy of the United States of America [R]. DOD, 2018.

略构想和战略规划的层面，尚未上升为行政层面可供实施的战略文件，且其聚焦点是国家战略而非军事领域。特朗普执政时期，围绕军事智能化发展，已经构建形成了“国家战略—国防战略—军种战略”三位一体、较为全面系统的战略体系。

在国家战略层面，特朗普总统于2019年2月11日签署了《维持美国在人工智能领域的领导地位》的行政命令①，提出了美国发展人工智能的政策目标，将“维护国家安全”与“促进科学发现”“提升经济竞争力”并列，作为确立美国在人工智能领域领导力的三个核心指标，体现了特朗普政府以国家战略方式牵引军事智能化发展的战略思路。

在国防战略层面，与国家战略同步，美国国防部于2019年2月12日发布《2018年度国防部人工智能战略概要》，明确了人工智能对于国防和军队建设的重要意义，提出了发展军事智能化的战略方针，将建设联合人工智能中心、提供支持关键任务的人工智能功能、增强态势感知和决策能力、提高操作设备的安全性和强化与技术领先的私营部门合作等作为国防部军事智能化战略的重点，并提出了具体的行动方案和途径方法②。

在军种战略层面，为响应和有效对接国防部人工智能战略，推进以人工智能为核心驱动力的新军事变革深入展开，美军各军种结合自身实际情况和军种建设的目标要求，分别提出了具有各自特色

① Executive Office of the President. Executive order on maintaining American leadership in Artificial Intelligence [EB/OL]. (2019-02-14) [2022-04-07]. https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence.

② Department of Defense. Summary of the 2018 Department of Defense AI Strategy [EB/OL]. (2012-02-11) [2022-04-07]. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

的人工智能发展战略。以美国海军和空军为例，美国海军司令部于2019年发布了《海军人工智能框架》，系统梳理了海军部门面临的挑战，揭示了人工智能技术在海军部门最重要的应用领域，探讨了如何将人工智能应用于重要的海军作战任务链接及相应的人员配备、组织要求和政策考虑因素等问题，并就如何帮助海军部门成为人工智能领域的适应性组织、如何有效而迅速地使用人工智能技术等问题提出对策建议[①]；美国空军于2019年发布了《2019年空军人工智能附录》，作为美国空军实施《空军2030科技战略》的框架，该附录详述了空军实施人工智能战略的背景和目的，同时，将降低技术准入门槛、维护数据战略资产、提供个性化人工智能解决方案、招募和培养专业人员、强化国际合作交流作为空军实施人工智能战略的重点领域。[②]

美国人工智能军事化战略设计的一个特色，即以军民一体化战略推动人工智能领域的协同创新。政府在推进人工智能领域的军事化应用进程中，采取了军民一体化发展方略，充分利用美国民用人工智能领域在研发资源、技术储备、人才培养、产品开发等方面的绝对领先优势，推进军民协同创新，并在战略规划、机构设置以及具体研发创新过程中充分体现。在战略规划方面，奥巴马政府在颁布《国家人工智能研发战略规划》时指出，美国人工智能研发战略计划并非专属军方的战略规划，而是在政府的全盘战略统筹下，与《国家大数据研发战略规划》《国家网络安全研究发展战略规划》《国家隐私研究战略》《国家纳米技术倡议战略计划》《国家计算战

① LEWIS L，VAVRICHEK D. An AI framework for the department of the Navy [R]. Center for Naval Analysis，2019.

② 刘晓娟，李志，王彤. 2019年美国空军人工智能附录：响应国防部人工智能战略[J]. 无人系统技术，2019，2（4）：59－61.

略倡议》《大脑神经技术创新研究倡议》和《国家机器人倡议》一起协同推进。[①] 在机构设置方面，为了统筹军地人工智能领域的研发管理，2016 年 5 月，奥巴马政府在国家科学与技术委员会中专门设置机器学习与人工智能分委会，其职能就是跨军地部门协调人工智能研究开发、资源配置以及监督管理等问题。[②] 以“通过推动创新型神经技术开展大脑研究”项目为例，该项目的研发汇聚了美国国防高级研究计划局、国家卫生研究院、霍华德·休斯医学研究中心和艾伦脑科学研究所等来自军方、政府以及民间的不同研究机构。

2. 打造多维一体的人工智能装备体系

谋求全方位的、可持续的军事竞争优势是美国发展军用人工智能技术的战略动因。作为“技术制胜”理念的长期践行者，美国将自主武器系统视为“改变战争规则”的颠覆性武器，试图通过发展自主武器系统来谋取未来军事竞争的可持续优势，不断完善政策制度设计，加快装备技术研发和作战应用的进程。早在 2010 年，美国国防部就在《四年防务评估报告》中指出：“现代战争的发展带来了技术上的重大进步。越来越多的无人系统应用于军事领域，展现出在精确性、持久性和自主性方面的广阔前景，伴随这些系统能力和数量的增长，我们必须确保我们使用武力的政策得以相应的推进。”[③] 在雄厚的科技、军事和财力支持下，美国在陆、海、空、天和网络空间的智能化武器装备建设方面居于全球领先地位，并打造

① National Science and Technology Council. The National Artificial Intelligence Research and Development Strategic Plan [EB/OL]. (2016 - 10 - 13) [2022 - 04 - 07]. https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf?ref=innovationpolicy.org.

② Obama White House. NSTC Committee on Technology [EB/OL]. (2012 - 02 - 11) [2022 - 04 - 07]. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/CoT%20Charter%202011.pdf.

③ Department of Defense. Quadrennial Defense Review Report [R]. DOD, 2010.

了一支数量规模庞大的智能化作战力量，已经建构了覆盖陆、海、空、天和网络空间的智能化装备体系。

在政策制度的牵引下，美军按照“全域威慑”和“全球快速投送”的目标要求，积极推进智能化武器装备的全域部署。陆军方面，2017 年，美国陆军构建了智能化的陆战目标瞄准和自动攻击系统，通过大数据和机器学习来提升陆战武器获取、识别和打击目标的速度；海军方面，美国海军不仅与美国国防高级研究计划局合作开发“水上猎手”无人水面舰艇，还于 2019 年公布《无人潜航器系统展望》，推动“虎鲸”等超大型无人潜航器项目，为机器人潜艇配备智能传感器和算法以自行执行多样化作战任务；空军方面，形成以“捕食者”和“收割者”无人机、X－47B 无人战斗机和“蜂群”式无人机集群攻击等为核心的智能化空中作战力量；网络空间方面，2009 年，美军成立了网络战司令部，并全力推动网络空间自主武器系统的研发，开发的网络武器多达 2000 种；太空方面，2018 年，特朗普政府下令组建的太空军成为美国武装力量的第六军种，同年，美国国家航空航天局宣布启动自主航天飞船研发项目，从而进一步加剧了外层空间的潜在战争风险。

3. 设置和完善军用人工智能管理机构

为保证军事智能化发展的战略规划能够顺利推进，美国政府和军方对现有机构设置进行了一系列调整和重组，覆盖统筹协调、研究开发和转化推广等方面的职能使命。

在统筹协调方面，2018 年，特朗普政府成立了国家人工智能安全委员会。这是一个在政府行政部门中专注于人工智能领域的独立委员会，由参议院军事委员会、众议院军事委员会和国防部任命的 11 名相关领域专家组成，负责从顶层规划和设计人工智能领域的整体发展方案，审查和评估人工智能技术进展和应用前景，构建和完

善人工智能技术安全风险监管机制，协调国际人工智能合作，从而全面满足维持美国在人工智能领域全球领导地位的需要。[①] 该委员会在推进军事智能化发展中扮演至关重要的角色，其内部下辖四个工作组，第二工作组的职责就是致力于打造和维持美国在国家安全领域的人工智能技术优势，负责对与军事相关的人工智能技术进行全面评估，并为增强美国军事竞争力提供咨询建议。[②]

在研究开发方面，由于人工智能领域的技术研发涉及多学科交叉创新，加之其军事应用范围覆盖情报侦察、信息操纵、指挥控制、火力打击、后勤保障、网络空间行动及半自主式和自主武器系统等诸多方面，为适应研发特点和提升研发效率，美国军方对原有国防创新部门进行改革重组。2018 年，国防部重组“国防创新试验单元”，成立新的“国防创新单元”，意在通过组织重构，打通阻碍人工智能技术流通的军地壁垒，强化军方与私营部门在人工智能研发方面的交流合作，同时，充分发挥市场的竞争性优势，为国防部提供更具性价比的人工智能技术解决方案。作为美国国防科技创新的引擎，美国国防高级研究计划局针对人工智能技术的前沿进展，成立名为“AI Next”的中长期研发项目组。该项目组在不限定具体研发领域的前提下，聚焦确保美国未来在人工智能领域的领导力，对具有前瞻性、先导性和颠覆性的第三代人工智能关键技术进行超前探索。类似“国防创新单元”、“AI Next”项目组的组织模式创新，使得人工智能研发机构更灵活、具有更强的自适应能力。

在转化推广方面，2018 年 6 月，美国国防部成立联合人工智能

① 阚天舒，张纪腾. 美国人工智能战略新动向及其全球影响［J］. 外交评论，2020，37（3）：136.

② Congressional Research Service. Artificial Intelligence and National Security［R］. Congress，2019.

中心，并在《2018 年度国防部人工智能战略概要》中明确指出：国防部建立联合人工智能中心，其目的就是加快人工智能技术交付应用的进程，扩大人工智能对国防部的影响力，使得国防部的人工智能研发进展能够同步转化为联合作战部队的优势。具体而言，联合人工智能中心负责快速交付具有人工智能功能的产品以完成关键国防任务，增强美军当前的军事实力优势，并根据任务需要加强未来人工智能的研发工作、运营成果、用户反馈和数据支持。[①] 由此可见，在美国国防部关于军事智能化的战略筹划中，联合人工智能中心发挥着关键的联络点作用。一方面，它对接人工智能技术孵化机构，负责将军方需求传递给研发部门；另一方面，它对接不同类型的军方用户，致力于将人工智能前沿技术向作战应用转化，通过打通研发到应用的“最后一公里”，不断扩展人工智能在整个国防部的影响力。

4. 颁布军用人工智能管理和规制准则

军事智能化发展可以带来超越对手的技术优势，也会引发一系列的伦理、法律和安全风险。2018 年 5 月，3100 多名谷歌员工依据“不作恶”的科技伦理准则，联名抵制军事智能化项目 Maven。[②] 这一震惊全球的科技伦理事件，使得特朗普政府意识到技术管理与规制的重要性。美国国防部在《2018 年度国防部人工智能战略概要》中，提出了“负责任的 AI 开发”理念，以及在合法、合乎道德的情况下使用人工智能技术的愿景和指导原则，强调国防部将通过加大

① Department of Defense. Summary of the 2018 Department of Defense AI Strategy [EB/OL]. (2012 - 02 - 11) [2022 - 04 - 07]. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

② FRISK A. What is Project Maven? The Pentagon AI project Google employees want out of [EB/OL]. (2018 - 04 - 05) [2022 - 04 - 07]. https://globalnews.ca/news/4125382/google-pentagon-ai-project-maven/.

投资来提升人工智能系统研发的合法性、稳定性、可靠性、安全性和透明性，探索解决军事智能化发展的数据标准和隐私保护等问题，提升对军用人工智能技术造成意外伤害和附带损害的态势感知能力。

为进一步将“负责任的 AI 开发”理念贯彻于军事智能化的应用实践，2019 年，美国国防部下属的国防创新委员会发布了《人工智能原则：国防部人工智能应用伦理的若干建议》，就美军如何合乎伦理和法律地使用人工智能技术，提出“责任、公平、可追踪、可靠、可控”五项指导原则，提出在国防部建立人工智能指导委员会、加强国防部人力资源培训计划、强化对人工智能新安全风险的研究、定义人工智能可靠性基准、加强人工智能测试和评估技术、制定人工智能风险管理方法等 12 条具体的咨询建议。该咨询报告于 2020 年 2 月为美国国防部所采纳。

值得注意的是，美国人工智能军事化战略是面向全球的外向型扩张战略，因此，特朗普政府在制定军用人工智能技术的管理规制原则时，不仅要考虑对内的应用安全性，而且要考虑对外巩固全球霸权的便利性。《2018 年度国防部人工智能战略概要》明确指出美军要成为全球军用人工智能伦理和安全的领导者，提出一套全球性的制度准则，并通过强大的对话渠道，将军事智能化应用的道德和安全愿景传递给尽可能广泛的国际受众。2019 年 8 月，美国商务部国家标准与技术研究院对外颁布《美国在人工智能领域的领导力：联邦政府参与制定人工智能技术标准的计划》，强调美国应当在全球人工智能管理规则和技术标准的制定中居于领导位置，特别要注意保证相关条约和标准制定的弹性和灵活性，规避因过度限制造成对美国人工智能发展战略的阻碍。

（二）动因分析

1. 应对非传统安全领域的威胁

应对非传统安全领域的威胁是美国推动军事智能化发展的重要考量。伴随全球化的不断深入发展，非传统安全威胁日益成为国际社会关注的焦点。冷战结束后，世界军事实力对比呈现美军一家独大的格局，美国面临的传统安全威胁相对下降，与此同时，其面临的非传统安全威胁相对上升。2018 年度美国《国家安全战略》指出："恐怖分子、跨国犯罪组织、网络黑客和其他恶意的非国家行为体对全球事务的大规模破坏能力不断增强。"在这一背景下，美国要凭借人工智能领域的领先优势来维护国家安全利益，就必然将传统安全与非传统安全两个领域面临的威胁等量齐观。在推进军事智能化发展过程中，美国立足于巩固军事领域的竞争优势，同时也积极探索如何利用人工智能应对日益严峻的非传统安全威胁。

军事智能化有助于提高美国反恐怖主义行动的效率。特朗普执政以后，积极探索人工智能和大数据技术在反恐领域的应用。军事智能化可以为美国反恐提供强有力的情报信息支持。利用人工智能技术在数据挖掘和数据分析领域的独特优势，可以对海量社交媒体数据和其他类型数据信息进行数据挖掘，对涉恐通信信号、语音记录、视频录像和社交网络中异常活跃的社交账号进行相关性分析，梳理其关注与联系对象，勾勒恐怖组织行动网络，为锁定和追捕恐怖犯罪嫌疑人、切断恐怖行动的指挥链路以及瘫痪恐怖组织提供情报支持。同时，人工智能技术还可以对潜在恐怖主义威胁进行预测、评估和早期预警。2019 年 1 月，美国国家情报总监办公室发布了

“利用机器增强智能战略”①，该战略通过将情报部门数据与人工智能算法相互结合，来增强反恐人员的推理预判能力，通过对恐怖袭击政治诱因、民族诱因、宗教诱因以及袭击时间与袭击位置分布等方面历史信息进行大数据分析，揭示其中可演化的规律，构建恐怖袭击风险预测模型，对未来可能发生恐怖袭击的概率进行预先评估。

军事智能化有利于维护美国网络安全。早在2016年，美国网络司令部司令罗杰斯上将在参议院军事委员会的讲话中表示，在网络空间中仅仅依靠人力系统是“一种失败的策略”。② 美国国防部认为，人工智能可以增强预测、识别和应对各种来源网络威胁的能力，加强对国土的防御，阻止敌对分子袭击破坏美国金融网络、电网、选举程序和医疗网络系统等国家关键基础设施的企图。③ 美国国会也认为，人工智能很有可能是推进网络空间攻防行动的关键技术。④ 在网络防御方面，传统的网络防御手段只能在静态环境下匹配过去已知的病毒，因此可以通过修改局部代码的方式绕开防御。将人工智能工具引入网络防御系统，能够检测更为广泛的网络入侵行动和系统异常情况，从而实现更为全面动态的网络防御。在网络攻击方面，2016年以来，美国国防高级研究计划局持续举办网络挑战大赛，邀请全球顶尖“黑客”展示人工智能的网络进攻潜力。人工智能工具

① Office of the Director of National Intelligence. The AIM Initiative: a strategy for augmenting intelligence using machines [EB/OL]. (2019-01-16) [2022-04-07]. https://www.dni.gov/index.php/newsroom/reports-publications/item/1940-the-aim-initiative-a-strategy-for-augmenting-intelligence-using-machines.

② Committee On Armed Services. Hearing to receive testimony on encryption and cyber matters [EB/OL]. (2016-09-13) [2022-04-07]. https://www.armed-services.senate.gov/imo/media/doc/16-68_09-13-16.pdf.

③ Department of Defense. Summary of the 2018 National Defense Strategy of the United States of America [R]. DOD, 2018.

④ Congressional Research Service. Artificial Intelligence and National Security [R]. Congress, 2019: 11.

能够快速自主地检测、评估和利用潜在的网络漏洞，达成网络入侵、木马植入、信息窃取、系统瘫痪等目的，人工智能算法的速度优势有助于美军掌控未来网络攻击行动的主动权。

军事智能化能够强化美国信息操控与反操控能力。植根于人工智能的“深度伪造”技术，以人工智能自动生成内容的方式实现对人类行为特征的模仿与伪造，生成更加逼真的图片、音频和视频文件。美国国会研究服务处指出，国家或其他非国家行为体可以将人工智能伪造或“深度伪造”部署为信息操纵的一部分。一方面，“深度伪造”技术可以制作针对竞争对手国家的虚假新闻报道，可用于在选举期间操纵舆论，削弱公众信任，引发群体事件，甚至制造社会恐慌。[①] 同时，还可以利用人工智能的信息操控能力伪造并创建完整的“数字生活档案”，将个人网络行为痕迹与购物历史、信用报告、社交网络以及订阅记录合并匹配，为军事人员和情报官员的海外行动提供便利。另一方面，人工智能也为防范敌对国家针对美国的恶意信息操控提供技术手段。美国国防高级研究计划局正在开展旨在打击“深度伪造”技术的“媒体鉴别分析”项目，该项目开发人工智能工具自动检测可疑信息操控行为的能力，分析和揭示各类信息操纵行为如何伪造详细信息及实现视觉媒体整体完整性的内在机理，为实现对“深度伪造”技术的智能化整体防控提供条件。

2. 巩固在全球人工智能领域的领导地位

就战略动因和目标而言，美国推进军事智能化发展既是为了捍卫美国的国家安全利益，也存在服从、服务于美国人工智能国家战略的内在考量。推动军事智能化发展是美国人工智能国家战略中不

① Congressional Research Service. Artificial Intelligence and National Security [R]. Congress, 2019: 28.

可或缺的一部分。2019 年 2 月 11 日，特朗普总统签署了《维持美国在人工智能领域的领导地位》的行政命令，将美国人工智能战略上升为国家战略；2019 年 2 月 12 日，美国国防部又出台了《2018 年度国防部人工智能战略概要》。这种前后相继的时间关系也充分说明了两者之间存在密不可分的内在联系。美国通过推动军事智能化发展来巩固美国在全球人工智能领域的领导地位，这一战略考量具体体现在以下三个方面。

一是回应竞争对手对人工智能领导权的挑战。人工智能领域是引领未来科技创新和社会变革的颠覆性领域。近年来，全球主要国家和国际性、区域性组织纷纷将发展人工智能作为推动经济发展、维护国家安全和提升国家综合竞争力的重要抓手，俄罗斯、欧盟、日本等国家和组织纷纷制定和发布了人工智能战略。军事智能化领域既是大国战略布局的密集领域，也是国际竞争的焦点领域。随着国际上军事智能化发展规模的不断增长，美国国会和国防部门对军事智能化领域日趋激烈的国际竞争表示忧虑。2017 年，在参议院特别情报委员会关于“全球威胁评估”的年度听证会上，科茨指出：“美国的竞争对手应用人工智能的能力将对我们产生深远而广泛的潜在影响，包括更容易遭受网络攻击、难以确定攻击源头、为外国武器和情报系统的进展提供便利、事故和相关责任风险日益上升，以及国内人工智能产业面临失业问题等。”[①] 在此背景下，美国推进军事智能化发展，抢占军事智能化领导权，本质上也是为了巩固美国在全球人工智能领域的领导地位。

二是领导人工智能治理标准的制定并向全球传递价值观。特朗

① Select Committee On Intelligence. Current and Projected National Security Threats tothe United States [EB/OL]. (2013 - 03 - 12) [2022 - 04 - 07]. https://www.congress.gov/113/chrg/CHRG - 113shrg82721/CHRG - 113shrg82721.pdf.

普政府在《维持美国在人工智能领域的领导地位》的行政命令中指出："美国在全球人工智能领域的领导地位，不仅体现在成为全球人工智能产品研发和部署应用方面的领导者，而且体现在制定并塑造全球人工智能领域的治理标准，确保全球人工智能领域的研发、应用符合美国的价值观和国家利益。"① 当前，世界主要国家围绕人工智能治理标准的竞争日趋激烈。以数据保护规则的制定为例。美国为充分发挥人工智能的技术优势，希望同其他国家共享数据资源，以实现数据在全球范围的自由流通和巩固经济、军事霸权。而欧盟国家基于自身利益和数据安全考虑，更加强调数据本地化和隐私保护，并于2018年5月颁布了《一般数据保护条例》。在此背景下，美国推进军事智能化发展也充分体现了对规则制定权的争夺。美国《2018年度国防部人工智能战略概要》不仅将"领导军事伦理和人工智能安全"作为军用人工智能发展的战略方针，而且将"提出并倡导一套全球性的军事人工智能准则"作为战略重点领域②，并积极推动联合国《特定常规武器公约》会议关于致命性自主武器系统军备控制标准的讨论，旨在制定符合美国军事利益的军备控制标准，减少军用人工智能技术研发测试和武器装备使用面临的潜在障碍，从而最大限度地发挥美军在智能化军备领域的先发优势，夺取在全球军用人工智能治理领域的先行话语权。

三是通过人工智能领域的军民一体化发展来实现整体领导力的

① The White House. Executive Order on Maintaining American Leadership in Artificial Intelligence [EB/OL]. (2019-02-11) [2022-04-07]. https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence.

② Department of Defense. Summary of the 2018 Department of Defense AI Strategy [EB/OL]. (2012-02-11) [2022-04-07]. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

提升。军事技术与民用技术之间存在双向的互动关系。一方面，美国在民用人工智能的研究开发、硬件设备、数据资源和人才储备等方面拥有绝对优势[①]，这为其实施军用人工智能战略奠定了坚实的基础；另一方面，推进军事智能化发展及其所产出的一系列成果，也可以反哺民用人工智能领域的发展。基于此，美国军事智能化发展的战略设计充分考虑军用人工智能对民用领域的反向推动，具体体现在以下三方面。一是美国军事智能化发展将一些军民通用技术作为战略重点领域。例如，图像识别算法技术，既可以识别网络视频中的猫，也可以识别经由叙利亚或阿富汗上空的无人驾驶飞机捕获的全动态视频中的恐怖活动。为了在这一通用技术领域占据领先优势，美国国防部成立了“算法战跨职能小组”，作为人工智能联合中心重点推动的第一个国家任务计划，并投入大量人力、物力和财力。二是美国的军事智能化发展强调营造开放、互动与共赢的军民合作环境。《2018 年度国防部人工智能战略概要》将“与商业界、学术界和国际盟友等合作伙伴合作”作为战略方针之一，明确指出“从研究到开发，在人工智能技术发展的每个阶段，维持牢固的军民伙伴关系都是必要的。国防部将与学术界和工业界合作，应对全球具有重大社会意义的挑战，并提供资金来吸引优秀的学者投身于与国防相关的长期研究；国防部将与全球开源社区合作并为之做出贡献，以推进新兴技术的研发和应用”。[②] 三是美国军事智能化发展的战略设计注重与民用领域互补。鉴于部分人工智能基础研究和应用基础

① 周琪，付随鑫. 美国人工智能的发展及政府发展战略［J］. 世界经济与政治，2020（6）：29.

② Department of Defense. Summary of the 2018 Department of Defense AI Strategy［EB/OL］.（2012 - 02 - 11）［2022 - 04 - 07］. https://media.defense.gov/2019/Feb/12/2002088963/ - 1/ - 1/1/SUMMARY - OF - DOD - AI - STRATEGY. PDF.

研究领域不具有明确的短期利润驱动，难以引起民用部门很大的兴趣，美国国防部提出发挥军事研究不计工本、不以赢利为目的的创新优势，以军事部门的前沿实验来孕育最具变革性的人工智能技术，再将趋于成熟的技术投入市场，通过这种军民一体化的互动机制来确保美国在全球人工智能领域的整体领导力不断提升。

三、其他国家和组织加速推进军用人工智能研发和应用

除美国外，其他国家和组织从维护安全利益和战略稳定性的考量出发，立足国情和军情，因地制宜地制定竞争策略，不断提升武器装备智能化水平，加速推进智能化武器装备的研发和应用进程，积极谋取军用人工智能领域的竞争优势。

（一）俄罗斯

俄罗斯确立了以军用机器人为核心、以点带面的竞争策略。俄罗斯是威胁美国在军用人工智能领域全球领导地位的重要挑战者之一。俄罗斯总统普京公开宣称："谁先掌握了人工智能，就会统治整个世界。"[①] 因国防预算和研发经费不足，俄罗斯难以在军用人工智能的各个领域与美国展开竞争，而是采取以重点突破带动整体推进的发展方略。从2008年开始，俄罗斯就在军事现代化改革中布局军用人工智能，逐步加大对自主武器系统的投资力度，确定将小型机

① DEAN J W. The US and global Artificial Intelligence arms race [EB/OL]. (2017-12-07) [2022-04-07]. https://www.veteranstoday.com/2017/12/07/neo-the-us-and-global-artificial-intelligence-arms-race/.

器人、机器人步兵编队和机器人坦克等作为俄军人工智能战略的重点领域。2019 年，俄罗斯总统普京签署命令，批准《2030 年前人工智能国家发展战略》，将人工智能发展提升至国家战略层面。[①] 俄军将以“平台－M”陆战机器人、“阿尔戈”战斗机器人和“铀－9”机器人坦克为代表的一系列智能化装备投入叙利亚和乌克兰战场，接受实战的检验。未来，俄罗斯将持续加大在军用机器人的研发和应用方面的投入力度。依据俄罗斯军事工业委员会设定的目标，到 2025 年，俄军 30% 以上的军事装备将实现机器人化。[②] 由此可见，俄军试图创造陆战机器人集群作战优势的意图明显。俄罗斯人工智能专家梅德韦杰夫甚至预言，2030 年俄罗斯军队将主要由自主机器人组成。[③]

（二）西欧主要军事强国

一方面，西欧主要军事强国高度重视军用人工智能发展，长期将研发重点放在军民两用的工业人工智能和机器人技术上。英国、法国、德国、瑞典和意大利等国都在致力于研发具有高度自主性的军用机器人系统。英国是重要的军用人工智能装备出口国，将无人机技术作为国防投资的重点方向，研制开发了“守望者”大型无人驾驶飞机、“忠诚僚机”无人机和具有完全自主模式的“塔拉尼斯”无人战斗机，其中，“守望者”无人机已经列装部队，用于英吉利海

① 李锡宁．俄制定人工智能国家发展战略［N］．中国国防报，2019－10－28（4）．

② SIMONITE T. For Superpowers, Artificial Intelligence fuels new global arms race［EB/OL］.（2017－09－08）［2022－04－07］. https://www.wired.com/story/for－superpowers－artificial－intelligence－fuels－new－global－arms－race/.

③ ATHERTON K D. Russian roboticist sees all－but－lethal autonomous weapon systems by 2030［EB/OL］.（2018－09－18）［2022－04－07］. https://www.c4isrnet.com/unmanned/2018/09/17/russian－roboticist－sees－all－but－lethal－autonomous－weapon－systems－by－2030/.

峡上空对越海难民的侦察和警戒。[①] 另一方面，西欧主要军事强国通过新的欧洲防卫基金，有效地汇集各国的专业技术人才，试图成为全球人工智能和自主武器系统军备竞赛中的主导者。西欧国家采取了分布发展与整体聚合相统一的竞争策略。在欧盟主导下，西欧国家围绕军用人工智能的合作日趋频繁。早在2012年，英国和法国就签署协议合作开发“暴风雨”未来战斗航空系统。2020年10月，意大利防务公司列奥纳多与英国皇家空军合作开展了名为Alvina的无人机“蜂群”作战试验。[②]

（三）非传统军事强国

非传统军事强国纷纷加快自主武器系统的引进、仿制和作战应用进程。为缩短与世界军事强国的差距，非传统军事强国将发展自主武器系统作为在军备领域“弯道超车”“换道超车”的重要战略举措。统计数据显示，截至2017年6月，全球掌握无人机技术的国家已经超过100个，掌握武装无人机技术的国家有16个，包括缅甸、哈萨克斯坦、土库曼斯坦、尼日利亚、沙特阿拉伯、阿拉伯联合酋长国、伊朗等非传统军事强国。[③] 印度的“奥拉”无人机项目以及韩国的SGR－A1“哨兵”机器人项目在自主化程度方面已经接近发达国家水平。

---

① Dronewars. An overview of Britain's drones and drone development projects [EB/OL]. (2021－03－01) [2022－04－07]. https://dronewars.net/british－drones－an－overview/.

② Dronewars. An overview of Britain's drones and drone development projects [EB/OL]. (2021－03－01) [2022－04－07]. https://dronewars.net/british－drones－an－overview/.

③ 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：113.

（四）非国家行为体

值得关注的是，虽然军用人工智能技术的研发和应用主要集中于少数军事强国，但是人工智能技术的高度军民一体化导致许多低成本的人工智能武器向非国家行为体扩散。形形色色的非国家行为体将自主武器系统应用于武装冲突、恐怖袭击等行动中，给国际安全环境带来新的挑战。目前，自主武器系统已经扩散至包括也门胡塞武装、“真主党”、“哈马斯”和“伊斯兰国”在内的诸多非国家行为体。[①]非国家行为体正尝试将无人机用作简易爆炸装置的打击载体，使致命性自主武器系统引发的暗杀袭击行动不断加剧。自主武器系统向非国家行为体的扩散，还导致国际上无人机“黑飞”事件频繁发生，严重干扰了国际社会正常的公共安全秩序。美国白宫、日本首相官邸、韩国青瓦台、法国巴黎市政府等主权国家敏感区域都曾遭遇无人机非法闯入。

① 沙瑞尔. 无人军队：自主武器与未来战争［M］. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：111.

第二章
人工智能军事化的国际安全影响

人工智能是军事变革的潜在推动者和力量“倍增器”，因此，各国大力发展军用人工智能技术，不断提升武器装备的自主程度，进而导致作战方式、力量编成乃至战争形态的智能化变革，人类日益丧失对战争的控制权。同时，人工智能领域的军事变革也会反向作用于国际政治，加剧世界范围内的军备安全困境，不可避免地引发新一轮全球军备竞赛，进而对国家间战略竞争态势、战略稳定性、战争文化观念和战争伦理产生持久而深远的影响。本章围绕战争形态、军备竞赛、战略稳定性和战争伦理四个方面，深度剖析和揭示人工智能军事化对全球安全格局的潜在影响，从而为我国积极应对国际安全风险提供启示。

一、激化战争失控的风险

人工智能作用于军事领域，极大地提升了武器装备的自主性，挑战了人类作为战争主体的核心地位。各类自主武器系统开始取代人类成为战争的主体，基于人类智慧的、以大平台为中心的传统作战方式，被基于算法的、去中心化的“蜂群”作战和“马赛克战争”方式所取代。人类日益从战争的组织者、参与者沦为战争的旁观者。在由人工智能主导的战争模式下，战争的不确定性和失控风险急剧增长。

（一）作战主体的“去人类化”

传统战争是“人类中心化”的战争，战争是以人类为唯一的主体而展开的。从冷兵器时代开始，人类操控武器作战是战争对抗模式的基本特征，武器必须与人紧密结合才能够形成战斗力。随着“冷兵器—火器—原子武器—精确制导武器”的演进，武器装备的打击半径和毁伤效能不断增长，但人类作为武器装备操控者的角色始终没有发生改变，人类作为战争主体的核心地位也始终未曾动摇，赢得战争的关键是人而非武器这一制胜机理也始终没有改变。

人工智能的军事化发展，促进武器装备智能化程度不断提升，对以人类为中心的传统战争模式造成了严峻的冲击和挑战。2001 年 2 月 16 日，美军使用“捕食者”无人机在内利斯空军基地成功发射了一枚 AGM－114C“地狱火”导弹并击中目标。美国历史学家马泽蒂指出，这是美国航空史上第一次在进行重大试飞时不需要担心飞

行员的生命安全问题，人类悄无声息地进入远程控制武器作战的时代。[①] 2002 年 11 月 4 日，在阿富汗战场之外的也门马利卜地区，美军使用“捕食者”无人机向恐怖组织目标发射精确制导炸弹，击毙“基地”组织重要头目阿尔哈特希。美国军事历史学家耶纳指出，此次袭击行动是第一次被广泛报道的由一架“捕食者”无人机单独实施的对一个高价值目标进行攻击的行动。[②] 集侦察、打击功能于一体的“捕食者”和“收割者”无人机，本质上已成为一种“平台无人、体系有人”的半自主武器系统。无人作战平台具有自主导航定位、情报侦察、火力打击和效果评估能力，人类从居于武器平台之中的操控者转变为处于平台之上的监督者，在必要的时候可以干预和接管武器系统的指挥与控制，人类在作战系统的地位由“人在回路之中”向“人在回路之上”转变。

当前，西方军事强国正在研制和开发的完全自主武器系统，将彻底颠覆以人类为中心主导战争的历史。未来的战争将是完全自主武器系统之间的对抗，完全自主武器系统将被大规模投入战场，成为战争的实际主体，人类从战争的参与者转变为战争的旁观者，人类在作战系统的地位将发生由“人在回路之上”向“人在回路之外”的颠覆性转变。在由机器主导的战争模式下，杀戮行为将不再受到人类的约束，战争将彻底失去控制。

（二）作战手段的自主化

人工智能的军事化发展，不仅将人类移出了战争的 OODA 环，

① 马泽蒂. 美利坚刀锋：首度揭开无人机与世界尽头的战争［M］. 王祖宁，王凌凌，美同，译. 北京：新世界出版社. 2014：84－85.

② 耶纳. 无人机攻击［M］. 刘春辉，赵希源，译. 北京：中国市场出版社. 2014：Ⅱ－Ⅲ.

导致作战主体的“去人类化”，而且在作战工具层面推动了武器装备自主性能的不断提升和完全自主武器系统的涌现。武器从纯粹意义上的战争手段日益演变为战争主体和战争手段的复合体。完全自主武器系统结合了视觉感知、语音、面部识别和智能决策等人工智能技术，以执行空中拦截、两栖地面攻击、远程打击和海上行动等军事任务，而无须人工干预和监督。[①] 完全自主武器系统的完全自主性体现在以下三个方面。

一是自主环境感知和目标识别。通过综合运用仿生视觉识别、基于智能神经网络的机器学习与深度学习、多传感器信息融合的战场态势评估等技术手段，自主武器系统不仅可以自主掌握战场气象、水文、地质等自然环境信息，而且可以感知攻击目标的力量分布、火力配置、武器装备和运动状态等情况，从而实现复杂战场环境下对攻击目标的精确感知和识别。以 2020 年 11 月发生的伊朗核科学家遇刺事件为例。伊朗革命卫队副司令法达维指出，美国和以色列在未部署任何地面人员的情况下，通过配置高清摄像头和人脸识别技术的自主遥控机枪，精准锁定并击毙了重重防卫下的伊朗核科学家法赫里扎德，而同车相距 25 厘米的法赫里扎德夫人却毫发无伤。[②] 智能化遥控和感知技术的精准度由此可窥一斑。近年来运用自主武器系统的刺杀行动日益频繁，成为威胁国际安全和地缘政治稳定的一大隐患。

二是自主研判和分析。随着信息技术的高速发展及其与战争系

① JOHNSON J. Artificial Intelligence, drone swarming and escalation risks in future warfare [J]. The RUSI Journal, 2020, 165 (2): 28.

② HAFEZI P, WILLIAMS D. Iran says "smart satellite - controlled machine gun" killed top nuclear scientist [EB/OL]. (2020 - 12 - 07) [2022 - 04 - 07]. https://www.nasdaq.com/articles/iran - says - smart - satellite - controlled - machine - gun - killed - top - nuclear - scientist - 2020 - 12 - 07 - 0.

统的深度融合，人类的反应速度越来越难以适应作战信息系统的运行速度。正如美军智能化作战专家丘奇博士所言："自然人的生理能力难以应对智能化作战所需的对海量数据的处理能力与决策速度。"① 依托大数据和人工智能技术，自主武器系统可以自主实现数据挖掘和数据分析。如美国国防部联合人工智能中心主导的 Maven 项目，可以加速处理来自无人机的大量视频片段，从而更好地锁定打击目标。② 自主武器系统还可以依据战场态势发展进行动态任务规划，提高作战响应的时敏性，通过自主分析缩短战场决策时间，加快作战速度。随着加速的作战决策及对决策的加速再响应，人类分析和理解战场态势的速度将跟不上智能机器运行的速度，难以有效履行智能系统监督者的职责。

三是自主决策和攻击。执行层级上的自主化是自主武器系统实现完全自主的关键环节。一方面是决策的自主化。自主武器系统可以依据数据信息分析进行作战推演，基于对战场的实时分析提供可行的行动路线方案，自主进行作战决策，从而提高战时决策的质量和速度。如美军利用特定算法对恐怖活动的发生概率进行评估，研究和预判恐怖袭击行动。另一方面是攻击的自主化。真正意义上的完全自主攻击，要求无人作战系统在不依赖外界指令和设备支持的情况下，在复杂作战环境中独立地做出开火决定。虽然对于目前大部分人类开发的自主武器系统而言，远程控制中心的操控员仍然掌握着开火的最后按钮，但是一些国家已经开始研制具备自主开火权

① CHURCH A M U. Air force world: RPA strikes still rising [J]. Air & Space Forces Magazine, 2013 (3): 96.

② PELLERIN C. Project Maven industry day pursues Artificial Intelligence for DoD challenges [EB/OL]. (2017-10-27) [2022-04-07]. https://www.airuniversity.af.edu/News/Display/Article/1356172/project-maven-industry-day-pursues-artificial-intelligence-for-dod-challenges/.

的完全自主武器。2020 年 3 月，一架隶属于利比亚政府军的卡古 -2 四轴飞行器向反对派武装利比亚“国民军”发起自杀性爆炸攻击，开启了无人机自主攻击人类的先河。英国《新科学家》杂志撰文指出，这次攻击展示了一种可以改变战争面貌的开火方式，人类通往自我毁灭的潘多拉魔盒已经开启。①

（三）作战方式的去中心化

近代以来，随着资本主义机器大工厂的蓬勃发展，人类战争进入了以大型机器平台为中心的平台对抗阶段，由此催生了以“大炮巨舰主义”为代表的大平台作战方式，以及闪击制胜论、坦克制胜论、装甲制胜论等平台制胜理论。“二战”期间，航空母舰成为海上作战力量的中心平台，依托舰载轰炸机群实施集约式杀伤，成为美日两国海军在太平洋战场上的主要作战方式。“二战”以后，信息化战争浪潮席卷全球，以海湾战争为代表的信息化战争，在强调制信息权的前提下，仍然高度重视大型平台对于战争制胜的关键性作用。配置精确制导武器的战斗机、航空母舰以及战略核潜艇，仍然是信息化作战网络的中心环节。

智能化战争时代，各类微型化、智能化作战平台如雨后春笋般不断涌现。较之传统的大型作战平台，小型作战平台具有无人化、低成本、零风险的作战特点和更高的战术灵活性，将挑战并日益取代以大机器平台为中心的传统作战方式。为实现作战方式的智能化转型，美国国防部提出“蜂群”作战理念，组建大量微型化战斗群，尝试分布式的军事力量部署，依靠小而多的数量优势对敌实施饱和

① HAMBLING D. Autonomous military drone may have attacked humans [J]. New Scientist, 2021, 250 (3337): 16.

攻击，通过高密度、高频率的持续攻击，破解对手既有的防御体系。美国国防部实施了 Perdix 计划，利用 F－16 和 F/A－18 战斗机部署微型无人机“蜂群”。微型无人机展示了先进的群体行为，如集体决策、自适应编队飞行和自我修复。整个“蜂群”作为一个集体有机体，共享一个分布式大脑进行决策。由于每架无人机与其他无人机进行交流和协作，因此“蜂群”中没有明确的领导者，像自然界中的群鸟一样相互适应。[①] Perdix 计划所倡导的“去平台化”和“去中心化”，其实质是要实现作战网络通路、节点的可替代性。依托“去中心化”的“蜂群”式部署，减少对大型作战平台的过度依赖，即便面临某些节点遭敌摧毁的状况，整体作战网络也能够规避被阻断或瘫痪的风险。

作战方式的去中心化，不仅体现为作战平台的可替代性，而且体现为作战系统和力量编成的动态整合，即根据作战环境、战略目标和任务要求的发展变化，展现出高度的灵活性和自适应性。2019 年，美国国防高级研究计划局战略技术办公室主任格雷森在一次会议上提出了“马赛克战争”，认为智能化战争制胜的关键在于有人和无人系统之间的适应性、协同性和互操作性。[②] 在“马赛克战争”模式下，作战力量必须具备模块化的组合功能，如“马赛克”积木一样自如地拆卸与拼搭，根据任务差异性利用智能化指挥系统进行自适应调整，快速应对开放战场环境下作战条件的复杂变化。随着作战方式智能化程度不断提高，战争的不可控性和出现意外风险的

① CLARKE C. US military deploys swarm of 3D printed “PERDIX” drones from fighter JETS [EB/OL]. (2017－04－03) [2022－04－07]. https://3dprintingindustry.com/news/us－military－deploys－swarm－3d－printed－perdix－drones－fighter－jets－109900/.

② SPIE. Timothy Grayson － Mosaic Warfare [EB/OL]. (2019－04－23) [2022－04－07]. https://spie.org/news/dcs－plenary－mosaic－warfare?SSO=1.

概率也越来越大。

二、加剧军备竞赛的安全困境

安全困境也称为安全两难，指一国为确保自身安全利益不受侵犯而采取的军备发展行为，可能被其他国家视作需要做出反应的军事威胁，降低包括敌对国家在内的其他国家的安全感，引发区域性甚至全球性的军备竞赛，从而导致该国陷入更加不安全的局面。冷战期间，美苏两个超级大国为了确保自身和本阵营的安全，都在极力将发展比对方强大的核军事力量作为战略目标，这种畸形、超饱和、持续升级的核军备竞赛，造成了愈演愈烈的军备竞赛安全困境。

军用人工智能领域既是大国战略布局的密集领域，也是国际竞争的焦点领域。为抢占新一轮军事变革的主动权，近年来，全球主要国家纷纷将发展人工智能作为推进国防和军队建设、维护国家安全和提升国家综合竞争力的重要抓手，美国、俄罗斯、日本等国家和国际组织纷纷制定和发布人工智能战略，不断加大研发投入，一场围绕人工智能领域的军备竞赛正在全球如火如荼地展开，由此引发的安全困境也日益显现。

（一）凸显军备发展的两难困境

作为一种在国家间安全互动中难以规避的问题，人工智能时代的军备发展同样面临日益严峻的安全困境。以美国为代表的军事强国大力推进军用人工智能技术的发展，打破了全球军备力量发展的既有格局，引发了世界其他国家的竞相跟随，传统军事强国和新兴国家纷纷参与其中，对军用人工智能领导者地位的竞争日趋白热化。随着各国因“预防式”的人工智能军备竞赛而不遗余力地投入，其

相互之间的安全感与信任度不断降低，军备发展的安全困境也不断加剧，具体体现在以下三个方面。

一是发展致命性自主武器系统的两难困境。以人工智能驱动的自主武器系统替代传统的人力系统，从而有效降低人力成本支出，这是奥巴马政府“第三次抵消战略”的战略核心和战略重点，也是美军维持全球军事竞争优势的重要基点。美国国会在《人工智能与国家安全》的咨询报告中指出，各类自主武器系统平台可以替代人类，从而促使更多的军人投身更为复杂和对认知要求更高的工作。用自主系统替代人类从事枯燥、肮脏或高度危险的任务，美军将从中受益良多。[①] 为了从政策层面进一步推动自主武器系统的研发、生产和使用，美国国防部于2012年11月颁布《国防部指令3000.09：武器系统中的自主性》，规定了美军研制、生产半自主武器和完全自主武器的通用准则，具体包括以下三个方面的准则：（1）寻的弹药等半自主武器；（2）以舰载“宙斯盾”武器系统为代表的防御型监督式自主武器；（3）非致命性和非动能杀伤的武器。[②] 这一指令打通了自主武器系统快速发展的道路，为一系列具体武器装备的发展打开了“绿灯”。在推动美军自主武器系统快速发展的同时，美军高层也充分意识到，放任武器装备自主性程度的不断提升，甚至是研制“激活后就无须操作人员进一步干预，可以自主选择和攻击目标”的完全自主武器系统，可能导致武器失控和战争无限升级的巨大风险，因此，必须对完全自主武器系统的研发采取审慎态度。美国国防部前副部长沃克曾经指出，美国在发展致命性自主武器系统方面

① Congressional Research Service. Artificial Intelligence and National Security [R]. Congress, 2019: 27.

② 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：97.

会努力坚守底线，不会将致命的权力下放给做出决定的机器。[①] 但是，美国发展自主武器系统的谨慎态度并不能阻止其他国家加速推动武器装备自主化程度的进程。据国际智库新美国研究中心披露，俄罗斯、英国、韩国等国家正在研制具备半自主/自主选择功能的智能化武器系统[②]，自主武器技术正在加速向包括恐怖组织在内的非政府组织扩散。自主武器系统的国际竞争日趋激烈的现实状况，也迫使美国对自主武器系统的发展政策留有余地，沃克指出，随着竞争的发展，如果美国的战略竞争对手更愿意将权力委派给机器，那么美国将不得不做出最有利于自己的竞争策略。[③] 沃克的这番言论，鲜活地反映了当前各国深陷人工智能军备竞赛的安全困境。

二是人工智能应用于网络、太空等新兴空间的两难困境。人工智能作为提升军事能力的“倍增器”，不仅在陆、海、空等传统军事空间得到广泛应用，而且日益向网络空间和外层空间渗透。美国等军事强国将人工智能应用于网络、太空等领域的军事化进程之中，其初衷无疑是进一步捍卫其在新兴空间的安全利益。然而，伴随人工智能与网络、太空军事化的结合不断深化，相应的军备竞赛和安全风险也随之加剧。

在网络空间，人工智能工具应用于互联网络，能够快速自主地检测、评估和利用潜在的网络漏洞，达成网络入侵、木马植入、信息窃取、系统瘫痪等目的，人工智能算法的速度优势有助于掌控未来网络攻击行动的主动权。美国国防部高度重视人工智能对于识别

① JOHNSON J. Artificial Intelligence, drone swarming and escalation risks in future warfare [J]. The RUSI Journal, 2020, 165 (2): 29.

② 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：111－130.

③ JOHNSON J. Artificial Intelligence, drone swarming and escalation risks in future warfare [J]. The RUSI Journal, 2020 , 165 (2): 30.

和应对网络威胁、加强国土防御，以及阻止敌对分子袭击破坏国家关键基础设施的能力。[①] 美国在网络空间的军事智能化进程，引发了包括俄罗斯在内的战略对手的警觉。2017 年，俄罗斯在名为 Zapad 的军事演习中展示了其网络武器的攻击能力。[②] 2020 年，美国联邦调查局披露：俄罗斯在 2020 年上半年投入了大量的金钱和精力，利用网络散布虚假信息和煽动舆论，其目的是破坏 2020 年美国选举。[③] 除俄罗斯以外，以色列 8200 部队、英国军情六处、德国网络和信息空间司令部、日本网络空间防卫队，都是这些国家开展智能化网络作战的重要力量。

外层空间因其维系国家安全的“高边疆”属性，成为当前全球人工智能军备竞争的重要维度。早在 1997 年，美国空军太空部助理部长霍尔就指出：“关于太空优势，我们拥有它，我们喜欢它，而且我们将继续保持这一地位。”[④] 自 2018 年美国国家航空航天局宣布启动自主航天飞船研发项目以来，各国围绕人工智能在外层空间的军事应用进程不断加快。2019 年 8 月，俄罗斯首次向国际空间站发射名为“费奥多”的人形机器人，以测试升级后的联盟号太空舱和联

① Department of Defense. Summary of the 2018 Department of Defense Artificial Intelligence strategy [EB/OL]. (2019 - 02 - 12) [2022 - 04 - 07]. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

② PAGANINI P. Zapad drills - Russia may have tested cyber weapons on Latvia [EB/OL]. (2017 - 10 - 06) [2022 - 04 - 07]. http://securityaffairs.co/wordpress/63918/cyber-warfare-2/zapad-drills.html.

③ EWING P. Wray: Russian interference heavy on agitation, less so state cyberattacks [EB/OL]. (2020 - 09 - 17) [2022 - 04 - 07]. https://www.npr.org/sections/live-coverage-threats-to-the-homeland-hearing/2020/09/17/913930966/wray-russian-interference-heavy-on-agitation-less-so-state-cyberattacks.

④ BONSOR K. How space wars will work [EB/OL]. (2000 - 11 - 01) [2022 - 04 - 07]. https://science.howstuffworks.com/space-war.htm.

盟号助推器的性能。[①] 俄罗斯航天局相关人士在推文中一针见血地指出，要征服太阳系的行星，必须在太空创建机器人殖民地[②]，由此展露其依托人工智能推进太空军事化的雄心壮志。此外，近年来，其他国家推动太空军事化和智能化的进程也不断加快。2020 年 5 月，日本防空部队成立了日本第一支名为“太空行动中队”的航天部队。2020 年 9 月，法国空军改名为法国空军和航天部队，强化其太空战专业力量建设。[③]

三是智能化核武器系统发展的两难困境。伴随人工智能军事应用的深度和广度不断拓展，其与核武器系统融合可能导致的安全困境，同样引发国际社会的关注。历史上，美苏等核大国利用人工智能开展核军备竞赛由来已久。冷战期间，美国就开始探索用人工智能来增强其核反击的可能性，并推出名为“生存适应性计划实验”的研究项目，试图利用当时的人工智能技术，使美国能够瞄准苏联的移动洲际弹道导弹发射装置。[④] 与之相对应，苏联军方担心美国第一次攻击会导致其失去对核武库的主动控制，在冷战期间设计部署了名为“周界”的自动化核反击武器系统，一旦苏军的核指挥控制系统遭遇美国核突袭后瘫痪，“周界”系统就会被激活，自动实施应

---

① HOWELL E. Russia Launches Humanoid Robot “Fyodor” to Space Station on Rare Soyuz Test Flight [EB/OL]. (2019 - 08 - 22) [2022 - 04 - 07]. https://www.space.com/russia-launches-humanoid-robot-on-soyuz-spacecraft.html.

② The Moscow Times. Russia's Humanoid Robot Says It Wants to Build Robot Colonies In Space [EB/OL]. (2019 - 09 - 19) [2022 - 04 - 07]. https://www.themoscowtimes.com/2019/09/19/russias-humanoid-robot-says-it-wants-to-build-robot-colonies-in-space-a67353.

③ China Military Online. Overview of international military conflicts and security situation in 2020 [EB/OL]. (2020 - 12 - 24) [2022 - 04 - 07]. http://www.81.cn/jwywpd/2020-12/24/content_9958111.htm.

④ 盖伊斯特，罗恩. 人工智能会怎样影响核战争风险？[J]. 国外社会科学文摘，2018 (8)：16.

对美国第一次核打击的报复行动。[①] 美苏之间的智能化核军备竞赛引发了一系列安全风险。例如，1983 年 9 月，云顶反射的阳光导致了苏联卫星的错误警报，苏联部署的“奥科”核预警系统自主检测到来自美国的核导弹袭击并发出虚假警报，造成 20 世纪 80 年代最大的核弹危机。[②] 当前，随着美国和俄罗斯在核打击和防御方面的智能化程度不断提升，类似冷战时期的核武器智能化安全困境可能会重演。美国军方和智库高度重视智能化作战平台在核打击和突防方面所具备的巨大军事前景。2011 年以来，美国海军持续开展 X－47B 隐形无人攻击机飞行试验。刘易斯和波斯托尔指出，利用隐形无人机攻击的突防能力，可以破坏敌对国家的核武器指挥控制系统、输送系统和关键核基础设施，从而进一步提升美国的战略威慑能力。[③] 美国蒙特利国际研究学院的卡伦伯恩等人认为，无人机“蜂群”技术有可能改善现有核武器运载系统并提供新颖的运载平台，人工智能技术还能够提高核打击的目标定位能力，强化其对抗传统防御体系的突防能力，促进新战术的发展，从而对全球核安全格局造成更大的不稳定因素。[④] 卡伦伯恩所指的“不稳定因素”主要是指可能引发俄罗斯等有核国家的恐慌情绪，从而加剧该领域的军备竞赛。作为对美方行为的回应，2015 年以来，俄罗斯开始研发名为“海洋多用途系统状态－6”的核动力水下无人潜航器，该无人潜航器携带

① THOMPSON N. Inside the apocalyptic soviet doomsday machine [EB/OL]. (2009－09－21) [2022－04－07]. https://www.wired.com/2009/09/mf－deadhand/.

② 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：1－2.

③ LEWIS G N, POSTOL T A. How US strategic antimissile defense could be made to work [J]. Bulletin of the Atomic Scientists, 2010, 66 (6): 8－24.

④ KALLENBORN Z, BLEEK P C. Swarming destruction: drone swarms and chemical, biological, radiological, and nuclear weapons [J]. The Nonproliferation Review, 2018, 25 (5－6): 529.

巨大的热核弹头，可以自动绕过反潜防御系统，对美国海岸实施毁灭性核打击。[①] 美国与俄罗斯围绕智能化核打击平台的竞争，可能进一步推动人工智能与核武器等大规模杀伤性武器系统的结合，从而诱发新一轮核军备竞赛，甚至可能导致冷战期间“恐怖平衡”式超饱和竞争的重演，使得核军控面临更加难以预测的前景。

（二）“能力－脆弱性”悖论日益显现

当前，全球围绕人工智能领域的军备竞赛存在显著的“能力－脆弱性”悖论。随着各国军用人工智能能力的提高，脆弱性问题也日益凸显，由此也导致当前围绕人工智能领域的国际性互动面临更加复杂尖锐的安全问题。

一方面，人工智能与军事系统的深度结合，使得各国在侦察预警、情报分析、指挥控制、毁伤评估和后勤保障等领域的军事能力增强，获得了在多维空间遂行智能化作战的能力。以美国为例，人工智能巩固了美国的军事竞争优势和全球军事霸权。据统计，美军在阿富汗和伊拉克战场上投入使用的无人机达到 7000 多架，超过 12000 个机器人承担作战、扫雷、防化、清障等各类军事任务[②]。与此同时，美国还非常注重在军用人工智能领域保持可持续的竞争优势。为此，美国国防部不仅实施了 X－47B 无人空天飞机、远程反舰导弹、快速轻量化自主武器和集群攻击作战武器等装备研发项目，面向未来智能化战争不断提升武器装备自主化程度，还与智库机构

① 盖伊斯特，罗恩. 人工智能会怎样影响核战争风险？［J］. 国外社会科学文摘，2018（8）：14.

② 沙瑞尔. 无人军队：自主武器与未来战争［M］. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：7.

合作，提出了“蜂群”作战、“算法战争”①、“马赛克战争”② 等一系列面向未来的智能化作战构想。美国在智能化作战能力方面的先发领跑，必然在西方军事强国之间引发“跟随效应”。各国加速推进人工智能技术向武器装备的转化进程，强化智能化武器装备研制、生产、开发和应用能力，导致围绕人工智能领域的军备竞赛愈演愈烈。

另一方面，各国军队对人工智能的依赖程度不断加深，脆弱性问题也随之出现。当前西方军事强国的人工智能军事系统严重依赖通信网络、计算机算法和数据库，导致其在面临网络入侵等挑战环境下的安全性和生存能力备受质疑。2011 年 12 月 5 日，伊朗军方俘获美军 RQ－170 隐形无人机的事例说明，即便战争一方拥有智能化武器装备和作战力量，另一方也仍然能够利用人工智能系统的脆弱性获取相对收益，通过网络瘫痪、算法篡改和数据破坏等方式实施非对称制衡，甚至发起先发制人的不对称攻击。

更为值得注意的是，“能力－脆弱性”悖论并不局限在常规武器方面，而是日益渗透到国家的战略核威慑力量之中。例如，在核指挥控制层面，美国《2018 年核态势评估》明确指出，国防部将寻求诸如机器学习之类的设计支持技术，以促进更有效、更快速的战略决策。③ 这种将人工智能技术引入战略威慑力量指挥控制体系的做法，虽然能够在一定程度上提升核力量运用的时敏性和效费比，但

① Utah Standard News. “Algorithmic Warfare”: DSD work unleashes AI on intel data [EB/OL]. (2017－05－01) [2022－04－07]. https://www.utahstandardnews.com/algorithmic－warfare－dsd－work－unleashes－ai－intel－data/.

② SPIE. Timothy Grayson － Mosaic Warfare [EB/OL]. (2019－04－23) [2022－04－07]. https://spie.org/news/dcs－plenary－mosaic－warfare?SSO=1.

③ JOHNSON J. Deterrence in the age of artificial intelligence & autonomy: a paradigm shift in nuclear deterrence theory and practice? [J]. Defense & Security Analysis, 2020, 36 (4): 422－448.

也会造成全球核大国之间围绕核指挥控制体系的整体智能化升级。英国智库国际战略研究所在研究报告中指出，当前，美国、俄罗斯、英国等有核国家的核指挥和控制系统越来越依赖人工智能程序，或者更准确地说，依赖专家系统和机器学习算法来增强信息流、态势感知和网络安全。[①] 针对这一现象，沙瑞尔表示，军方使用人工智能系统的重大风险在于人工智能参与发动核打击的决策，可能导致核突袭、核战争的爆发具有更强的不可预测性。[②]

三、威胁全球战略稳定性

人工智能军事应用不仅会对人类战争形态和全球军备发展造成颠覆性影响，还可能导致国家间战略力量平衡的重新调整，全球战略稳定性将不可避免地受到人工智能领域军备竞赛的影响。当前，军用人工智能的发展及其引发的全球性军备竞赛，可能在首攻稳定性、危机稳定性和军备竞赛稳定性三个方面对全球战略稳定性构成挑战。

（一）威胁首攻稳定性

首攻稳定性又称第一次打击稳定性。在战略稳定性理论体系中，首攻稳定性取决于先发制人与后发制人之间效费比的差别。如果先发制人打击较后发制人打击而言具有无可比拟的优势，那么实施先

① FITZPATRICK M. Artificial Intelligence and nuclear command and control [J]. Survival, 2019, 61 (3): 82.

② 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王娉，龙坤，译. 北京：世界知识出版社，2019：356.

发制人打击的意愿就会增强，首攻稳定性就会降低。[①] 一种新技术或武器系统是否显著影响首攻稳定性，既取决于该技术或系统的内在能力对现有攻防格局的影响，也取决于它如何部署及其运作方式与结构。[②] 从成本与收益的视角分析，人工智能的低成本和高收益，有利于强化技术领先国家的先发打击优势，从而降低第一次打击稳定性和战略威慑的有效性。

首先，人工智能可以大幅度地降低发起第一次打击的成本。打击成本及其潜在的风险直接影响先发攻击的决心。依据战略稳定性理论，一个国家使用某项新技术以显著改变首攻稳定性的能力，与技术研发和部署的成本成正比，技术成本越高，先发打击的意愿越弱，首攻稳定性越高。就研制成本而言，较之传统战略武器高昂的成本，各类人工智能化作战平台不需要前沿的、昂贵的尖端科技，甚至可以通过非军用设备生产大批量廉价的、一次性的打击载具。一架四轴无人机只需要 25 美元，而一台价值 35 美元的小型 Raspberry Pi 计算机可以运行先进的人工智能系统，足以在战斗模拟中击败空军战斗机飞行员。[③] 就部署成本而言，利用远程遥控的人工智能化作战平台实施军事打击，不需要向交战区域派驻一兵一卒，也不需要承受因士兵阵亡而引发的国内外舆论压力和政治风险。同时，考虑到人工智能攻击可以通过匿名方式来实现，因此也无须承担遭遇报复的风险成本。另外，借助 3D 打印技术和实施“蜂群”作战方式，人工智能的打击成本优势得以最大化发挥。借助廉价的

① 李彬. 军备控制理论与分析 [M]. 北京：国防工业出版社，2006：79.

② CHYBA C F. New Technologies & Strategic Stability [J]. Daedalus , 2020, 149 (2): 150 - 170.

③ HANER J, GARCIA D. The Artificial Intelligence arms race: trends and world leaders in Autonomous Weapons development [J]. Global Policy, 2019, 10 (3): 334.

3D 打印技术，可以轻而易举地在未来联网战场上部署数以亿计的微型无人机。一架孤立的无人机不太可能对传统武器系统构成重大威胁，但成千上万架由人工智能增强的无人机可能使传统防御系统不堪重负，从而进一步激发先发打击的意愿。

其次，人工智能有利于增加第一次打击的收益。一方面，人工智能打击平台可以用来穿透复杂的多层防空系统，直接打击敌首脑机关、核武器系统和国家关键基础设施等高价值目标，其巨大军事收益有助于激发军事强国实施先发制人军事打击的意愿；另一方面，利用自主武器进行先发制人的第一次打击，看似是战术层面的攻击行动，实际上可能产生战略层面的巨大收益。人工智能对于第一次打击收益的提升作用可以通过特定的部署方式来加强。在奥特曼等人看来，自主武器必然会为先发制人的攻击提供更强有力的激励。[①] 这种激励同样体现在高收益方面。近年来，国际上出现的“无人机袭击沙特油田”“无人机袭击委内瑞拉总统”“无人机刺杀伊朗军方领导人”“自动机枪刺杀伊朗核科学家”等一系列危机事件，证明技术先发国家只需使用远程操控智能化武器平台，就可以发起对敌对国家政府首脑、军方统帅或国家关键基础设施的“定点清除”式攻击，可能达成“首发制穴”的打击效果。更为值得关注的是，自主武器系统存在与核武器等战略武器相互融合的潜在趋势。依据战略稳定性理论，一国是否采取先发制人的打击策略，关键在于打击行动是否能够给予对手足够的伤害，实质性削弱甚至消除对手的反击能力。刘易斯和波斯托尔等学者指出，可以利用隐形无人机攻击敌对国家的核武器输送系统、指挥控制系统和关键基础设施组件，

① ALTMANN J, SAUER F. Autonomous weapon systems and strategic stability [J]. Survival, 2017, 59 (5): 130 - 132.

从而提升美国的弹道导弹防御能力。[①] 自主武器技术与核武器攻防技术深度融合并应用于核打击领域，可能给战略对手带来难以估量的政治、经济与军事损失。

最后，自主武器系统独特的作战方式可以打破现有的攻防格局，导致“进攻占优”的攻防不对称局面。攻防格局是影响第一次打击稳定性的重要因素。当攻防格局处于“进攻占优”的状态时，先发制人的攻击成本较低，更容易引起战争的爆发；当攻防格局处于“防御占优”的状态时，会使侵略他国的成本变高，从而增强第一次打击稳定性。自主武器系统的发展，为打破现有的防御体系提供了新的技术手段。以伊朗核科学家遇刺事件为例，在伊朗安全部门看来，刺杀者之所以能够深入伊朗境内成功狙杀目标，并使得伊朗方面严密的安保措施形同虚设，关键在于采用了卫星远程操控自动机枪的作战方式。[②] 2019 年 9 月，沙特油田遭遇也门胡塞武装无人机袭击时，防空系统竟然毫无反应。据媒体报道，虽然沙特阿拉伯已在先进的防空系统上投入巨资，但是大多数防空系统旨在防御传统威胁，并且装备不足，无法应对无人机等非对称空中威胁。[③] 这些案例表明，现有安防设施并不足以应对自主武器系统的密集攻击。基于此，近年来，欧美军事强国大力发展大量高速、群体模式的无人机“蜂群”攻击，将其作为突破现有防御体系的关键潜在力量。一旦以无人机“蜂群”为代表的自主武器系统成为突破现有防御体系

---

① LEWIS G N, POSTOL T A. How US strategic antimissile defense could be made to work [J]. Bulletin of the Atomic Scientists, 2010, 66 (6): 8 - 24.

② CBS NEWS. Nuclear scientist killed by satellite - controlled machine gun that "zoomed in" on his face, Iran says [EB/OL]. (2020 - 11 - 07) [2022 - 04 - 07]. https://www.cbsnews.com/news/mohsen - fakhrizadeh - iran - scientist - killed - satellite - controlled - machine - gun/.

③ Tim Lister. Attack on Saudi oil field a game - changer in Gulf confrontation [EB/OL]. (2019 - 09 - 15) [2022 - 04 - 07]. https://edition.cnn.com/2019/09/15/middleeast/saudi - oil - attack - lister - analysis - intl/index.html.

的利器，处于危机状态的国家施行先发制人打击的意愿就会增强，危机稳定性也就相应降低。

（二）威胁危机稳定性

危机稳定性是战略稳定性的重要组成部分。冷战期间，学界试图用先发制人打击（第一次打击）的后果与后发制人打击（报复打击）的后果之间的差别，来评估核威慑的有效性和衡量核打击的稳定性。如果先发制人与后发制人的后果差别小，则先发制人进行核打击的意愿低，危机稳定性高；反之，如果先发制人与后发制人的后果差别大，则先发制人的意愿就很强烈，危机稳定性低，核战争爆发的可能性高。[①] 危机稳定意味着即使在危机环境中（可能包括常规战争或即将发生的核战争），国家也不会优先采取使用核武器或其他使危机升级的行动。冷战结束后，危机稳定性也被用来衡量处于危机状态下的双方或多方行为对于控制危机所产生的影响。如果某个行为导致危机得到控制并最终化解，则危机稳定性高；反之，如果导致危机加剧和升级，则危机稳定性低。伴随人工智能技术的大规模军事应用，人工智能系统取代人类掌握战争的指挥和控制权力，人工智能系统的复杂性和脆弱性进一步增强了危机爆发与升级的不确定性、不可控性，从而降低了危机稳定性。

首先，人工智能带来危机预警方式的颠覆性转变，基于算法的危机预警机制日益取代传统人力主导的危机预警机制，从而增加了意外战争发生的风险。在战略稳定性理论中，意外战争是指无意或错误的战争升级，具体是指基于误解或错误情报的升级及由指挥和

① 李彬. 军备控制理论与分析［M］. 北京：国防工业出版社，2006：79.

控制故障导致的升级。[①] 冷战期间，美苏两国曾多次收到涉及核打击的危机信息，但由于人类在危机决策中掌握决定权，最终避免了大规模核战争的爆发。军用人工智能在战争风险评估与预测领域的应用，带来了作战指挥和决策领域的变革。美国国防部提出“算法战”构想并构建“算法战跨职能小组”，其本质就是利用计算机算法替代人类来处理海量数据信息，通过部署由大数据分析、网络功能和人工智能增强技术支持的智能化预警系统，实现情报分析和作战决策的智能化。[②] 这种基于算法的危机预警模式，将人工智能应用程序集成到预警系统（尤其是核预警系统）中，虽然可以缩短情报分析处理时间，提升危机状态下预警和决策的速度，但也将人类从“侦察—指挥—控制—打击—评估”的作战链条中移出。人工智能算法运行的复杂性已经超越了人类的认知能力，使得指挥员难以进行有效的控制和干预。在人机双重核查的危机管控机制无法实现的情况下，一旦使用复杂神经网络的预警系统产生意外的输出和指令，其创建者也可能难以识别或分辨，只能任由危机不断加速升级，甚至导致意外战争的爆发。以 2007 年 10 月美军在南非军事演习的意外事件为例。具有高度自主性的防空系统突然出现故障，导致美军参演防空火炮整体失控，虽然该防空系统的操控员多次尝试关闭火炮自动发射装置，但火炮已经被计算机系统接管而造成关闭指令失效，只能等待防空火炮系统中装弹机内的高爆炮弹自行发射完毕，该意外事故也造成美军多名参演官兵死亡或重伤。[③]

---

① CHYBA C F. New Technologies & Strategic Stability [J]. Daedalus , 2020, 149 (2): 151.

② Utah Standard News. “Algorithmic Warfare”: DSD work unleashes AI on intel data [EB/OL]. (2017-05-01) [2022-04-07]. https://www.utahstandardnews.com/algorithmic-warfare-dsd-work-unleashes-ai-intel-data/

③ 宋广收. 如何确保自主武器安全可控 [N]. 解放军报, 2020-08-18 (7).

其次，在危机爆发后，人工智能赋予自主武器系统对武装冲突进行“自主升级”的权力，导致人类“终止战争”的机制失去效力。构建“终止战争”的机制是确保对危机升级实行有效管控的重要方式。在传统战争模式下，交战各方政府和军方领导人通过外交磋商、签署停战协议或递交降书等方式掌握“终止战争”的控制权。当前全球范围内展开的人工智能领域军备竞赛，加速推动各类自主武器系统取代人类日益成为战争的主体，这类系统一旦激活后就无须操作人员进一步干预，使得危机存在“无限升级”的可能性，进而导致战争规模、毁伤程度和持续时间不受人类控制的局面。以“震网”病毒攻击为例，作为应用于网络空间的自主武器系统，它不仅成功入侵了伊朗布舍尔核电站的工业控制系统，造成用于提取浓缩铀的离心机严重损毁，而且造成伊朗境内67.6%的个人电脑和全球超过10万个计算机终端感染病毒。病毒还扩散到美国、中国、印度、巴基斯坦、沙特阿拉伯、印度尼西亚、阿塞拜疆和澳大利亚等众多国家。[①] 一种针对特定目标的蠕虫病毒便可引发全球范围内的网络危机，人工智能军事应用引发的危机升级风险由此可以窥见一斑。在战略安全领域，大量的自主武器系统可能应用于对核指挥系统和打击平台的攻击，从而对未来战争中的核安全和升级产生重大的战略影响。[②]

最后，人工智能使得匿名攻击高价值目标成为可能，从而进一步增加了危机爆发和升级的风险。冷战时期，美苏两个超级大国之间的战略稳定是建立在恐怖的核均衡之上，双方在危机环境下之所

① 张煌，杜雁芸. 人工智能军事化发展态势及其安全影响［J］. 外交评论，2022（3）：116.

② JOHNSON J. Artificial Intelligence, drone swarming and escalation risks in future warfare［J］. The RUSI Journal，2020，165（2）：26.

以没有爆发核战争，根源在于一方的核打击将导致对方毁灭性的核报复。由人工智能技术驱动的远程遥控攻击和机器自主攻击，极大地增加了攻击者与攻击目标之间的距离，隐蔽了攻击行为与攻击者之间的一一对应关系，导致对于锁定幕后肇事者的归因困境。近年来，国际社会出现以无人机自杀式攻击、卫星遥控半自主机枪等方式刺杀国家政要、首席核科学家等重要人物的匿名攻击事件，导致了地区安全局势的恶化。这种只享有使用武力权利而无须承担责任的攻击方式，无疑会导致攻击者有恃无恐地滥用武力，进而对危机稳定性构成挑战。在伊朗核科学家法赫里扎德遇刺后，联合国发言人就中东地区可能出现的安全危机发出警告，并呼吁中东地区各国保持克制，避免危机爆发和升级的风险。[①] 由此可见，自主武器暗杀袭击事件日益频繁，其对危机稳定性造成的负面影响已经逐渐成为共识，自主武器系统日益从多个维度对既有的危机稳定性产生新的冲击。作为全球人工智能军事化的领跑者，美国已经开始探讨使用智能化的隐形运载系统发动核攻击和打击对方核设施的可能性，从而引发了其他有核国家的警惕和不安。[②] 一旦以高超声速飞行器、无人机“蜂群”攻击为代表的匿名攻击方式成为突破现有导弹防御体系的利器，就会进一步激发危机状态下国家间实施先发制人打击的意愿。与之相对应，被打击国也可能对所有的潜在攻击者采取更高等级的报复措施，进而导致危机的不断升级，甚至核战争的爆发。

① China Daily. Iran says top nuclear scientist assassinated near capital [EB/OL]. (2020-11-28) [2022-04-07]. https://www.chinadailyhk.com/article/150822.

② ALTMANN J, SAUER F. Autonomous weapon systems and strategic stability [J]. Survival, 2017, 59 (5): 131.

（三）威胁军备竞赛稳定性

依据军备控制理论，“一种军备行为是否会引起对手的反应并导致军备竞赛，这个状态被称为军备竞赛稳定性”。[①] 因此，军备控制的主要任务是调整军备格局，由此降低潜在对手之间爆发核战争的可能性，降低双方出现军备竞赛的可能性。[②] 当前，正如美国人工智能专家奥莫洪德罗所指出的：“自动武器军备竞赛已经在进行。它们可以更快、更有效、更不可预测地做出反应。”[③] 西方军事强国在人工智能领域加快军备力量发展，开启了军备竞赛的“行动—反应”循环，导致全球军事强国围绕人工智能的军备竞赛稳定性进一步降低。人工智能技术具有军民通用性、不透明性和风险认知的模糊性，加之发展军用人工智能的成本低、障碍小，导致军用人工智能技术的迅速转移和扩散，加剧了全球军备力量平衡的不确定性和脆弱性，也导致当前围绕人工智能的军备控制面临重重困境。当前，国际上发展致命性自主武器的军备行为，导致全球致命性自主武器的军备竞赛愈演愈烈，主要从以下三个方面影响军备竞赛的稳定性。

首先，人工智能技术的军民一体化特征降低了扩散的成本。一方面，人工智能技术的军民通用性降低了其军事应用的成本。较之核武器，自主武器系统较低的研发难度为其向全球范围扩散提供了便利。奥特曼等人指出：“核武器的研发甚至到现在仍需要举国之力来实现，必须依靠集中的政治、军事努力来进行‘全国一盘棋’的统筹协调。与诸如核武器之类的大规模杀伤性武器装备相比，自主

① 李彬. 军备控制理论与分析［M］. 北京：国防工业出版社，2006：83.

② 李彬. 军备控制理论与分析［M］. 北京：国防工业出版社，2006：83－84.

③ MARKOFF J. Fearing bombs that can pick whom to kill［N］. New York Times，2014：A1－A18.

武器系统更加容易获得。”[①] 从技术层面来看，自主武器系统的核心部件在于自主控制系统，也就是在于软件而非硬件。由人工智能驱动的自主武器系统不需要昂贵的、难以获取的原材料或受到严格管制的绝密技术。正如奥特曼所指出的：“自主武器系统不需要矿石、离心机、高速保险丝或其他相对‘异国情调’的组件以机密的方式组装和测试。因此，核武器技术扩散可以受到控制，自主武器系统却难以进行监管。”[②] 美国国会研究处指出，许多人工智能工具用于民用领域，也可适用于军用武器系统，这些人工智能工具在非机密的互联网网站上被广泛共享，使得主要军事大国和非国家行为体都可以访问这些工具。[③] 民用领域人工智能产品的商业化运营和批量化生产，导致无人机等智能化运载平台转化为军事应用的成本迅速下降，加之机器学习的使用和普及，使得低成本的智能机器人技术市场蓬勃发展，其转为军用的难度亦随之下降。自主武器系统在全球范围内的快速扩散增加了军备控制的难度。另一方面，人工智能产品的军民一体化特征增加其出口管控的难度。较之军备控制条约，国际间出口管制协议不具有法律层面的强制性，导致其对国际军品贸易的约束力受到限制。更值得关注的是，自主武器向非国家行为体的流动，可能导致军备竞赛稳定性面临更为严峻的威胁。通过国际贸易渠道，技术后发国家、非国家行为体可以轻易地掌握军地通用的人工智能技术、设备和文件资料，并将其应用于致命性自主武器的研制与开发，进而增加这些系统对国际安全构成的潜在风险。

① ALTMANN J, SAUER F. Autonomous weapon systems and strategic stability [J]. Survival, 2017, 59 (5): 126.

② ALTMANN J, SAUER F. Autonomous weapon systems and strategic stability [J]. Survival, 2017, 59 (5): 126.

③ Congressional Research Service. Artificial Intelligence and National Security [R]. Congress, 2019: 33.

其次，军用人工智能技术的不透明性增加了核查难度。军备发展的不透明性是诱发军备竞赛的重要因素。在军备控制领域，潜在的战略对手之间可以通过提升军备发展的透明度，来抵消针对对方军备发展而产生的夸大的威胁感知，促成双方采取更为克制的军备行为，构建军备控制的相互信任机制和提升军备控制稳定性。在军备发展的透明性方面，军用人工智能技术与核武器截然不同。核武器的开发和部署需要获得的稀有资源（如用来进行可控裂变的铀）、难以隐藏的浓缩基础设施，以及对武器和运载系统的明显测试，这些都可以被专业机构检测到。相比之下，军用人工智能的发展通常被认为更为隐蔽。与网络武器一样，隐藏军事人工智能开发的“基础设施”和武器测试可能会更容易，从而削弱了核查是否符合军备控制协议的能力。[①] 自主武器技术发展的不透明性，导致自主武器的军备竞赛更容易升级。历史上，由不透明性造成的安全困境是引发军备竞赛的重要原因。在冷战时期，美苏双方有能力衡量彼此的核军备能力，却无法判断彼此的核军备意图，只能通过旷日持久的核军备竞赛来维持“恐怖的平衡”。[②] 对于发展自主武器技术的国家而言，不但难以洞察其他国家的军备意图，而且难以准确评估其他国家发展自主武器的军备能力。由于自主武器与半自主武器的区别在于软件而非硬件，这意味着很难从技术层面核查特定国家是否在发展真正意义上的完全自主武器。自主武器系统所拥有机器学习能力的不确定性，加之参战规则的可变性，进一步强化了自主武器系统作战能力的不透明性。军备意图的不透明性与军备能力的不透明性双重叠加，导致更多的不信任与误判行为的出现。由于担心被其他

① MASS M. How viable is international arms control for military artificial intelligence? Three lessons from nuclear weapons [J]. Contemporary Security Policy, 2019, 40 (3): 289 - 290.

② 谢林. 冲突的战略 [M]. 王永雄，译. 北京：华夏出版社，2019：246.

国家赶超，各国可能会在自主武器系统研发上投入更多的资金，导致自主武器的军备竞赛的不稳定性进一步提升。

最后，国际社会对于人工智能军事应用价值和应用风险的认知存在较大差异，可能迟滞军备控制的进程。国际社会对自主武器应用后果的认知差异，影响军备控制的进程。军备控制的成功前景还取决于国际社会对于以下两方面因素认知的综合考量：一方面是对武器军事应用价值的认知，另一方面是对武器使用所造成严重后果的认知。具体而言，如果人们认识到一种武器的应用价值有限但使用会造成可怕后果，那么针对其实施军备控制的成功可能性较大；反之，一旦某种武器能够起到决定战争进程的关键作用，即便其军事应用可能带来大规模杀伤性后果，实施军备控制的成功可能性也很小。对于人工智能武器而言，一方面，国际社会普遍认可人工智能技术具有较高的军事应用价值和广阔的应用前景。特别是在无人机等较为成熟的军事技术领域，人工智能先发国家对于推进军备控制的意愿较弱。奥特曼指出，考虑到美国、英国和以色列对武装无人机在中东执行军事任务的价值认同越来越高，它们不可能很快接受在这一领域的军备控制。[①] 另一方面，对于军用人工智能技术特别是致命性自主武器潜在应用风险的认知，各国存在较大的分歧，国际社会难以达成共识。部分技术后发国家，以及联合国、红十字国际委员会等非政府组织认为，致命性自主武器系统一旦应用于战争将导致战争彻底失控，因此必须禁止自主武器系统的研发。2018 年 11 月，联合国秘书长古特雷斯在纪念第一次世界大战结束 100 周年的巴黎和平论坛上指出："一旦致命性自主武器系统具有夺取人类生

① ALTMANN J. Arms control for armed uninhabited vehicles: an ethical issue [J]. Ethics and Information Technology, 2013, 15: 145.

命的权力，这在政治上是不可接受的，在道德上是令人反感的，应受到国际法的禁止。”[①] 与之截然相反的是，美俄等国从巩固其技术先发优势的立场出发，提出自主武器系统具有人类无可比拟的优越性。阿金认为，自主武器系统无须考虑自身安全，不受主观情绪左右，具有更快的信息处理速度，因而能够更好地履行道德规约，使战争变得更人性化。[②] 2018 年 3 月，美国在向联合国政府专家组会议提交的文件中指出：“致命性自主武器系统领域的新兴技术可以提高国家运用武力时降低平民伤亡风险的能力……较之压制或禁止致命性自主武器系统领域新兴技术的做法，美国选择鼓励相关领域技术创新。”[③] 以美国为代表的技术先发国家政府和智库机构认为，研发和使用致命性自主武器系统虽然存在一定风险，但可以通过计算机算法植入伦理道德规约等手段来规避潜在的安全风险，以确保其近乎完美地遵循战争伦理和法律要求。美国国防部国防创新委员会提出人工智能应用伦理的建议，就美军如何合乎伦理和法律地使用人工智能技术提出基本原则。美国道德与新兴技术研究所翁布鲁等人也指出，植入道德和法律的机器人可用来代替人类参与战争，并且通过扩大其应用范围，将人类从战争中解放出来。[④] 当前，两种观点针锋相对，很难在短期内达成共识与一致。

由此可见，国际社会对于军用人工智能发展前景与应用风险的认知确实存在巨大的分歧，这也是迟滞其军备控制进程的重要因素。

---

① 张煌，杜雁芸. 全球自主武器系统的发展及其对战略稳定性的影响 [J]. 国际论坛，2021，23 (3)：79.

② ARKIN R C. The case for ethical autonomy in unmanned systems [J]. Journal of Military Ethics，2010，9 (4)：333 - 334.

③ 徐能武，葛鸿昌. 致命性自主武器系统及军控思考 [J]. 现代国际关系，2018 (7)：56.

④ UMBRELLO S. The future of war：could lethal autonomous weapons make conflict more ethical? [J]. AI & Society，2020，35 (1)：273.

虽然早在2014年联合国在瑞士日内瓦召开的《特定常规武器公约》会议上就开始探讨自主武器系统的军备控制问题，但是直到2017年11月才召开第一次致命性自主武器系统的政府专家组会议，且主要职责是界定致命性自主武器的内涵与范畴，并未商讨签署多边军控条约的可能性。①

四、挑战战争伦理规约

各国军队自主武器系统的自主性程度越来越高，使用频次越来越密集，由此产生的战争伦理危机和人道主义灾难事件也越来越多。其中，美国无人机反恐造成的伦理危机和人道主义灾难，成为全球舆论和国际社会的热点话题。“9·11”事件以来，美军以“反恐”为名，长期在阿富汗、巴基斯坦、伊拉克、索马里、也门和叙利亚等主权国家领空范围内执行无人机“定点清除”行动，殃及大量无辜平民。美军无人机反恐的正义性与合法性问题，已经引发了伦理学界和法学界的广泛讨论。2020年1月3日，美军在伊拉克使用无人机击毙了伊朗“圣城旅”指挥官苏莱曼尼，这是美国首次针对主权国家的重要军事领导人实施“定点清除”行动，此次刺杀事件在正义性方面面临尖锐的争议。2020年11月27日发生的伊朗核科学家遇刺事件，又使得自主武器系统匿名攻击问题引发国际关注。自主武器系统的隐秘性，使得强权国家领导人相信依靠自主武器系统实施远程打击不需要承担战争责任且毫无风险可言，也使得这种匿名攻击的战争伦理后果比以前更难评估。由于各军事强国竞相研发、

① ICRC. Expert meeting on lethal autonomous weapons systems [EB/OL]. (2017-11-15) [2022-04-07]. https://www.icrc.org/en/document/expert-meeting-lethal-autonomous-weapons-systems.

部署和使用自主武器系统，智能化时代的军事对抗日益呈现出“去人类决策”“去人类中心化”的发展趋势，人类生存的安全环境面临前所未有的自我毁灭风险。自主武器系统正在取代人类的战争主体地位，掌握控制战争的权力，进而对传统的战争伦理观念构成强有力的冲击与挑战。

人类对于战争正义性的研究最早可以追溯到古希腊罗马时期，柏拉图、亚里士多德、西塞罗等人类早期思想家关于国家、政治和军事的学说中就已经蕴含了战争正义的基本理念。他们关于实现战争正义的具体观念虽然各不相同，但对于正义战争的基本态度、认知和立场大致相同，即正义的战争以实现和平、秩序和公益为根本目的。

在正义战争理论的发展过程中，基督教哲学家奥古斯丁提出将“至德至善”作为战争伦理评判的基本依据。欧洲中世纪著名的神学家、哲学家阿奎那第一次明确提出了关于发动正义战争的三个基本前提，即战争发动者必须是主权性质的权威和战争不是私人争斗，战争必须有正当而充分的开战理由，战争必须具有惩恶扬善的正当目的和意图。荷兰法学家格劳秀斯的著作《战争与和平法》第一次系统地阐述了正义战争理论的两个重要体系：“正义的战争”和“战争中的正义”。格劳秀斯提出的正义战争思想体系，对近代以来的国际法理论和实践产生了巨大而深远的影响。20 世纪上半叶，格劳秀斯所主张的正义战争观被进一步融入实证主义的条约法，并集中体现于《国际联盟盟约》《白里安－凯洛格公约》《联合国宪章》等国际公约之中。①

① 周桂银. 国际政治中的外交、战争与伦理［M］. 南京：南京大学出版社. 2018：276－285.

20世纪中叶以来，在经历第二次世界大战、朝鲜战争和越南战争等大量的战争实践之后，围绕正义战争理论的研究进一步深入。1977年，美国当代政治哲学家、哈佛大学政治学教授沃尔泽的著作《正义与非正义战争：通过历史实例的道德论证》，在广泛案例分析的基础上，系统考察了人类战争实践中出现的大量“道德两难”问题，结束了战争伦理规约长期以来的模糊、笼统状态，构建了涵盖开战正义、交战正义和战后正义的正义战争理论体系，并提出了一系列具体的战争伦理原则和标准，被公认为当前评判人类战争和低烈度军事行动正义性的权威理论依据。

由沃尔泽等人构建的经典正义战争理论体系，主要是以20世纪80年代以前的战争作为分析材料。20世纪90年代以来，人类战争形态从机械化向信息化、智能化转变，各类无人作战平台和作战手段得到大规模应用，信息化武器技术创新成为推动战争伦理演进最为活跃的要素之一。一方面，信息科学技术与武器装备推陈出新，对现有战争伦理体系构成持续挑战，促使人类反思信息化战争行为，制定新的伦理规约；另一方面，战争伦理规约一旦形成，并在国际范围内取得共识，就会在一定时期内对信息化武器装备的研发、应用及扩散产生道德约束。

正义战争理论认为，战争的正义性应当体现在目的、手段和责任三个方面，必须实现开战正义、交战正义和战后正义的统一。发展军用人工智能对战争伦理规约的挑战主要体现在开战正义、交战正义和战后正义三个方面。

（一）对开战正义的挑战

开战正义是实现战争正义的起始点，主要是探讨国家诉诸战争权利的正义性问题。依据正义战争理论及相关国际法规定，影响开

战正义的伦理原则主要包括合法权威、正当理由、正当目的、成功的可能性、相称性和最后手段六个方面。[①] 合法权威，即私人不得宣战，必须以主权国家的宪法和国际法为依据；正当理由，即正义战争的理由必须是明晰而准确的是非判断，必须以能证明确实受到侵害为前提条件；正当目的，从长远来看，开战是为了和平、秩序和正义，而非出于私人的贪欲或仇恨；成功的可能性，即不应发动成功希望渺茫的战争；相称性，即发动一场战争的道德成本不应同预期的道德收益相去甚远；最后手段，即战争始终是不得已而为之的最后选择。

当前，对于自主武器系统引发的战争伦理风险探讨，主要是将其作为一种战争手段，分析其在交战过程中的手段正义性问题，以及体现战后正义的责任分配问题。也有少数学者关注到战争手段对于战争目的的反向影响，如德国达姆施塔特工业大学的马蒂亚斯·恩格勒特等人认为，自主武器系统在作战手段中的道德应用，也是将价值观和行为准则输入智能化作战平台的控制系统。[②] 自主武器系统的军事应用，特别是智能化情报分析和作战辅助决策系统的出现，将机器意志渗透到开战决策之中，反向影响了战争决策者看待战争的价值观念与行动操守，从而对开战正义的伦理原则构成冲击与挑战。具体而言，自主武器系统对开战正义的冲击与挑战，主要体现在以下三方面。

1. 对正当理由原则的挑战

在正义战争理论体系中，正当理由是维系开战正义的首要原则。

---

① 左高山. 战争镜像与伦理话语 [M]. 长沙：湖南大学出版社，2008：31.

② ENGLERT M, SIEBERT S, ZIEGLER M. Logical limitations to machine ethics with consequences to lethal autonomous weapons [EB/OL]. (2014 - 11 - 11) [2022 - 04 - 07]. https://arxiv.org/abs/1411.2842.

正义战争理论对正当理由原则有严格的限定，即唯一正当的开战理由是针对外敌侵略进行自卫。自主武器系统的军事应用，冲击了自卫作为开战正当理由的唯一性。

一方面，当前自主武器系统作战应用所展现的低成本和高效率，导致技术先发国家不断弱化正当理由的伦理标准。较之传统作战平台，目前各类半自主和完全自主武器系统不仅具有平台无人、造价低廉等优势，而且可以通过匿名攻击规避战争责任。自主武器系统在经济、政治、人员安全方面的成本优势，使其广泛应用于高技术局部战争、反恐作战等行动中。以美国的无人机反恐为例。沃尔泽指出："较之其他杀伤形式，无人机容易太多了，以至于让我们产生不安。它是既诱人又危险的技术手段。它使我们的敌人比以往更脆弱，而我们的士兵没有任何风险。……但技术太好，使用的标准就会逐步放松。"① "察打一体"的武装无人机，不仅强化了西方军事强国打赢"反恐战争"的手段优势，也相应降低了其发动战争的物质成本与道德成本，导致少数霸权国家弱化约束开战正义的道德准则，以预先防卫的名义，滥用开战权利。

另一方面，未来完全自主武器系统的军事应用，将使机器意志渗透到开战决策之中，导致维系开战正义的正当理由原则面临更严峻的挑战。伴随军事智能化在全球如火如荼地展开，自主武器系统参与战争决策的程度日益深化。早在1990年，美军作战实验室进行实战化仿真的推演，形成了以"左勾拳"为核心的低风险、低伤亡的作战方案，成为促成美国老布什政府做出海湾战争决策的关键依

① WALZER M. Just & unjust targeted killing & drone warfare [EB/OL]. (2016-09-01) [2022-04-07]. https://direct.mit.edu/daed/article/145/4/12/27115/Just-amp-Unjust-Targeted-Killing-amp-Drone-Warfare.

据。[①] “9·11”事件以来，美国等智能化技术先发国家构建了“棱镜”等智能化大数据分析平台，[②] 运用大数据挖掘、模拟仿真、兵棋推演等技术手段预判可能存在的国家安全威胁，以此作为发动“先发制人”军事行动的凭据。当前，以美军为首的北约国家在“反恐战争”中，广泛应用人工智能技术预判恐怖威胁，对过往恐怖袭击时间、空间、人员、装备及政治诱因、宗教诱因、民族诱因等历史信息进行关联分析，预测未来可能发生的恐怖袭击事件和锁定恐怖嫌疑人。美军针对潜在恐怖嫌疑人实施先发制人的“特征攻击”，即一旦被自主武器系统认定具有恐怖分子的某些行为特征，就可以在未经国会批准的情况下实施“斩首”行动。[③]

未来，智能化战争决策系统自主能力的深度发展，可能进一步推动开战决策模式由机器辅助决策向机器代替人决策的颠覆性转变，由此也引发对开战理由公开性与透明性的挑战。依据正义战争理论，正当的开战理由不仅是合法的，而且应当是公开透明、为公众所理解的。自主武器系统一旦掌握开战决策权力，将导致公开的开战程序日益被隐秘的算法、模型和运算程序所取代，以各类算法、模型为基础的军事威胁评估系统被视作严格保护的机密，这些不透明、不接受质疑、难以被理解的算法和模型为战争决策提供关键依据，国际社会和普通民众无法了解决策程序的具体运行规则，很难对其决策结果的正义性提出反馈和质疑。

---

① 俞中明，杨旭. 海湾战争地面决战的运筹分析 [C]. 中国系统工程学会军事系统工程委员会第七届年会论文集，1995：54－57.

② KHUDAYER A Y, ABDULSALAM R M. Impact of NSA－PRISM to national information security strategy & policy [J]. International Journal of Information and Communication Technology Research, 2014, 4 (1): 7.

③ BYMAN D. Why drones work, the case for Washington's weapon of choice [J]. Foreign Affairs, 2013 (4): 32－43.

2. 对合法权威原则的挑战

自主武器系统对开战正义的挑战，同样体现为对合法权威原则的冲击。随着军用人工智能在作战指挥控制领域的应用，人工智能系统参与战争指挥决策的程度日益深化。21 世纪以来，美国构建了以“棱镜”为代表的智能化情报搜集分析平台和智能化条件下的作战实验室，运用大数据和人工智能等技术手段预判潜在的军事安全威胁。人们倾向于依赖计算机控制系统来判断的决策方式，已经严重损害了对自主武器系统的人为控制，因此必须对智能系统的可靠性保持谨慎，以免对操作系统过分信任。[①] 正如美国宾夕法尼亚大学约翰逊所指出的：“自主武器系统引发了诸多伦理问题，但最核心的问题是将决定人类生命的权力由人转交给机器。”[②] 作为开战正义的重要原则，合法权威原则规定战争是国家与国家之间的行为，主权国家和联合国是发动正义战争的合法权威。自主武器系统的军事应用，使得各类智能化作战平台取代人类进行战争筹划与决策成为可能，不仅颠覆了传统作战模式，也冲击了合法权威原则的道义基础。当前，美军在作战决策中对于人工智能系统分析报告的依赖度与日俱增，完全自主武器系统可以独立掌握开火权，也在无形中将本应归属于主权国家的开战权利让渡给了人工智能系统，导致存在国家意志被机器意志操控的潜在风险，进而冲击了合法权威原则的道义基础。

更为值得关注的是，自主武器系统一旦掌握自主开火权，成为发动战争的“合法权威”，就可能增加“意外战争”爆发的风险。

---

① FLEISCHMAN W M. Just say "no!" to lethal autonomous robotic weapons [J]. Journal of Information, Communication and Ethics in Society, 2015, 13 (3-4): 304.

② JOHNSON A M, AXINN S. The morality of autonomous robots [J]. Journal of Military Ethics, 2013, 12 (2): 129.

在传统的危机预警模式下，“人在系统的回路之中”，人类既是危机反应的最终决策者，也是危机处理最后的“保险装置”。沙瑞尔指出：“人类并不完美，但他们能够同情对手并考虑全局。与人类不同，自主武器无法理解其行动的后果，也无法在战争的边缘悬崖勒马。”① 冷战期间，美苏双方多次收到错误的核攻击信息，得益于人类的理性判断和人机双重检查机制，最终选择了克制而非升级的应对策略，避免了核战争的意外爆发。自主武器系统可能带来危机预警模式的革命性变化。自主武器系统拥有“一组可以转化为算法的判断规则”。② 在危机处理规程中，自主武器系统通过目标识别提取匹配算法规则的有效信息，自主做出具体的分析和判断。这种算法主导的自主预警模式，虽然在一定程度上提升了国家危机预警的速度和效率，但是，由于将人类从危机预警的系统回路中移出，人机双重检查制度和重新考虑机制无法实现。哈斯认为：“使用自主武器的国家会在危机方程式中引入一个超出其直接控制范围的变量。”③ 缺失人类干预造成的不可控性与算法运行的不可控性双重叠加，进一步增加了危机的不稳定性和“意外战争”的风险。

3. 对最后手段原则的挑战

在正义战争理论中，战争作为解决国际争端的一种手段，受到严苛的条件约束，即战争必须是政治、外交手段穷尽之后的最后选择。这是由于战争作为国之大事，关乎国家、民族的生死存亡，战

---

① 沙瑞尔. 无人军队：自主武器与未来战争［M］. 朱启超，王娉，龙坤，译. 北京：世界知识出版社，2019：357.

② Lin P, BEKEY G, ABNEY K. Autonomous military robotics: risk, ethics, and design［EB/OL］.（2008－12－20）［2022－04－07］. https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1001&context=phil_fac.

③ 沙瑞尔. 无人军队：自主武器与未来战争［M］. 朱启超，王娉，龙坤，译. 北京：世界知识出版社，2019：345.

争失败的沉重代价，导致战争手段的运用只能是不得已而为之。人工智能技术具备人力无可比拟的运算优势，使用人工智能技术可以对战争行为进行实时动态感知、模拟推演和推理决策，并对战争结果进行预测评估，导致战争尚未开始，胜负已在掌握之中。2018年，美军成立联合人工智能中心，其首要工作就是研究应对关键作战任务的人工智能能力，运用人工智能技术改进战场态势感知和作战决策效率，从而更好地实现人机融合、智能决策。[①] 一旦战争对于美国等军事强国而言毫无风险，它们就可以罔顾最后手段原则的约束，肆意发动针对其他国家的非正义战争，使国际安全环境陷入更加动荡的局面。近年来，国际上自主武器系统攻击事件频发，对于美国等军事强国而言，低成本、高效率且可以保证攻击方绝对安全的各类自主武器系统，已经将战争变成零风险的“机器”，“最后手段”蜕变为“最优手段”。概而言之，自主武器系统对最后手段原则的挑战具体体现在三个方面。

首先，自主武器系统挑战了最后手段原则成立的前提条件。费尔巴哈在《宗教的本质》中指出：“人类的恐惧多源自未知。”战争作为人类处理纷争最后手段的前提，也源自交战各方对战争结果的未知。在信息技术不发达的古代，人们无法准确预知战争的胜负，处于“战争迷雾”笼罩下的政府对战争的决策必然是慎之又慎。人工智能技术以及自主武器系统的军事应用，使得战争的结果在一定程度上从“未知”变为“已知”。人工智能科技领域的突破性进展和军事应用，催生了战争工程、兵棋推演以及作战实验室等一系列战争预判与评估手段。利用上述科技领域的最新进展，推演战争的进程并预测战争的结果，成为智能化战争的重要特征。借助人工智

① 李明海. 推演智能化战争面貌［N］. 解放军报，2019－04－02（7）.

能技术的风险评估、推演和研判功能，少数霸权国家在预知战争结果的情况下，将遥远的、不确定的军事威胁评估为确定的、迫在眉睫的战争前景，作为滥用“预先防卫权”和发动“预防性战争”的现实凭据，同时，利用自主武器系统攻击的低成本、零风险、隐秘性等优势，逃避“尸袋综合征”带来的国际、国内舆论压力，从而在轻易逾越战争门槛的同时占据道义制高点。

其次，自主武器系统打破了逐渐升级的最后手段的使用模式。对于战争与政治的关系，克劳塞维茨在《战争论》中有过精辟的论述：“战争是政治通过另一种手段的继续。”从时间关系上来看，这句话也说明政治手段与战争手段之间存在一个前后接续的关系。传统的战争往往经历由照会、声明、磋商、谈判和斡旋等低成本政治外交手段，再向军事渗透、介入、威慑、干预、打击等高成本、高风险战争手段逐渐升级的过程。各类自主武器系统的涌现，以其“无人化”“非接触”的打击优势，打破了逐渐升级的最后手段的使用模式。特朗普政府使用无人机击毙伊朗“圣城旅”指挥官苏莱曼尼，反映了在解决国际争端中，较之暴力程度更低但程序烦琐、时间冗长的政治、外交手段，能够一锤定音且毫无风险的自主武器系统显然更受到政治家的青睐。

最后，自主武器系统进一步挑战了维系最后手段原则的均衡性基础。最后手段原则成立的基础，在于敌对双方力量总体上处于均势状态，从而使他们审慎思考未来战争的残酷性和取得战争胜利的惨重代价。正如美国内战时期著名将领罗伯特·李所言：“战争残酷是件好事，否则人们真的会爱上战争。”奥特曼等学者指出，自主武

器系统的军事应用，进一步打破了国际军事竞争的战略稳定性。[①] 自主武器系统实现人与武器的实质性分离，极大地提升作战效率，同时使己方战争伤亡无限趋近于零。美军正在实施的“第三次抵消战略”，就是试图利用人工智能技术和自主武器系统的质量优势，抵消战略对手在传统信息化作战力量上的数量优势。未来，具备人工智能的无人军团，可能成为世界一流军队的“标配”，从而进一步拉大国际间军力差距，在国际军事博弈中塑造强者恒强、弱者愈弱的格局。

（二）对交战正义的挑战

交战正义又称战争手段正义，主要指交战各方必须严格限制战争手段的使用，使之遵循战争法和国际人道主义原则的约束。随着各国加大对自主武器系统的研发投入，其在交战过程中的自主性不断提升。例如，英国军方研发的“雷神”无人机，设置有主动攻击目标的可选模式，可在战场上自主搜寻和攻击敌对目标；[②] 韩国军方部署的 SGR－A1“哨兵”机器人，可以使用弱光摄像头和模式识别软件锁定敌人，并在必要的时候自主决定开火射击；[③] 俄罗斯军方装备的“平台－M”陆战机器人，可以在智能化系统控制下自主搜集情报、搜索和摧毁固定及移动目标。[④] 上述具有高度自主性能的武器系统引发了国际社会的关注，自主武器系统能否在交战过程中遵循

① ALTMANN J, SAUER F. Autonomous weapon systems and strategic stability [J]. Survival, 2017, 59 (5): 131.

② 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：120.

③ KUMAGAI J. A robotic sentry for Korea's demilitarized zone [J]. IEEE Spectrum, 2007, 44 (3): 16－17.

④ 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：125.

战争伦理规范，成为当前国内外学者争议的热点。一旦完全不受约束的自主武器系统取代人类成为战争唯一主体，必将挑战和冲击维系交战正义的基本伦理原则，可能导致人类生存环境面临前所未有的自我毁灭风险。军用人工智能的发展，主要从区分原则和相称性原则两个方面，挑战了战争手段正义的伦理规约。

1. 对区分原则的挑战

区分原则是体现交战正义的一项基本原则。在交战正义层面，接受和发展自主武器系统的最大障碍，就在于它们目前无法可靠地区分合法和非法目标。[①] 作为交战正义的重要体现，区分原则要求交战各方严格区分军事人员与平民，严格区分军用设施与民用设施，不应滥用武力，不能对民众赖以生活的基础设施和生存环境造成不必要的破坏。正如沃尔泽所指出的："在任何情况下都不能攻击非战斗员。永远不能将其作为军事行动的目标。"[②] 区分原则强调交战各方必须对攻击目标严格加以区分，以达成对平民特别是老弱妇孺、俘虏和丧失战斗能力者等弱势群体的保护。

人工智能的军事应用，导致区分方式的革命性转变，基于人力的道德区分被基于人工智能的算法区分取代。在传统战争模式下，区分目标由人类完成，自主武器系统利用算法取代人力主导的区分方式，即通过计算机编程的方法，将区分规则转化为自主武器系统能够理解的一组算法。算法区分的实质是通过目标识别提取匹配算法规则的特征信息，只能在相对简单的环境下实现对军服、制式武器等明显标识物的识别与区分，难以应对不断变化的环境、多样化

① UMBRELLO S, TORRES P. The future of war: could lethal autonomous weapons make conflict more ethical? [J]. AI & Society, 2020, 35 (1): 273 - 282.

② 沃尔泽. 正义与非正义战争：通过历史实例的道德论证 [M]. 任辉献，译. 南京：江苏人民出版社，2008：168.

的武器及复杂目标的特征。虽然在复杂多变的战争环境下，人类也是容易犯错的，但人类可以运用判断力识别出微妙的信号。例如，人类可以识别敌人何时想投降、埋葬死者或协助非战斗人员。[①] 较之人类，自主武器系统尚不具备这样的识别和区分能力，难以将需要人类主观意识判断的道德依据转化为机器能理解和执行的算法。如果盲目依靠计算机算法去区分，就可能导致对无辜平民生命的任意剥夺。

2018 年 5 月，3100 多名谷歌员工联名抵制军用无人作战项目 Maven。[②] 这一事件成为震惊全球的科技伦理事件，集中反映了学界对于算法区分模式潜在风险的忧虑。乌姆布雷洛一针见血地指出，机器算法区分存在区分目标的盲区，无法可靠地区分合法与非法目标。[③] 在他看来，并非所有的区分标准都可以转化为算法植入机器之中。当前，军服、武器装备等一系列明显标识可以设置区分规则，然而，在近期爆发的几场高技术局部战争和“反恐战争”中，为应对北约国家的军事打击优势，技术后发国家采取寓军于民的方式，其军事目标缺乏明显的视觉标识，给自主武器系统执行区分任务带来巨大挑战。卡尔皮等人进一步指出，自主武器系统难以胜任涉及目标主观情绪区分的任务。[④] 对于自主武器系统而言，区分投降信号的真伪比区分军民困难得多，因为传递投降信号的方式、方法和途

① LARK M. The future of killing: ethical and legal implications of fully autonomous weapon systems [J]. Salus Journal, 2017, 5 (1): 62.

② FRISK A. What is Project Maven? The Pentagon AI project Google employees want out of [EB/OL]. (2018 - 04 - 05) [2022 - 04 - 07]. https://globalnews.ca/news/4125382/google - pentagon - ai - project - maven/.

③ UMBRELLO S. The future of war: could lethal autonomous weapons make conflict more ethical? [J]. AI & society, 2019: 1 - 9.

④ KARPPI T, BHLEN M, GRANATA Y. Killer Robots as cultural techniques [J]. International Journal of Cultural Studies, 2018, 21 (2): 107 - 123.

径呈现巨大的多样性。人类对于投降信号的识别源自自身的经验、阅历和主观感受，这些难以通过编程方式转化为可供机器区分操作的算法。此外，钱铖等人认为，自主武器系统在执行区分原则方面还存在跨文化环境障碍，美军在无人机反恐作战中，将是否携带管制武器作为区分恐怖分子的重要特征，忽视了中亚地区游牧民族成年男性有随身佩戴刀具的文化传统。[①] 事实上，任何区分操作都是在具体的地域环境、文化环境和宗教环境中进行的，将自主武器系统置于跨文化环境的陌生战场，可能造成识别区分操作的失误。

在笔者看来，即便技术的发展使得上述区分盲区不复存在，以自主武器系统取代人类来区分目标，在道义上仍是难以接受的，因为它违背了确立区分原则的初衷。区分原则作为战争伦理规约的重要内容，它所体现的根本精神就是对人类生命权利的尊重。沃尔泽在《正义与非正义战争：通过历史实例的道德论证》中提及的“裸兵”案例启示我们，无论对于目标的识别多么契合区分原则，都需要在战争决策中留有道义的余地，以彰显敬畏生命、怜悯弱者的人性光辉。一旦赋予自主武器系统决定敌人生死的权力，就从根本上抹去人性在战场上的存在感，进而可能带来巨大伦理风险。导致伦理风险的原因，不是算法区分的运行错误，而是算法区分剥夺了人类决定同类生死的权力，违背了“珍视生命、同情弱者”的社会价值观念。在一个由自主武器系统主宰杀戮的社会，人类缺乏对杀戮行为的道德负担，其结果必然使杀戮变得毫无人性。

2. 对相称性原则的挑战

在正义战争理论中，相称性原则是区分原则的重要补充。相称

① 钱铖，石斌.“雄蜂”的阴影：美国无人机作战对当代战争伦理的挑战［J］. 世界经济与政治，2013（8）：96.

性原则强调在针对合法目标进行攻击时，必须视目标和作战环境的具体情况选择使用相称的作战方式和暴力手段，避免过度杀伤和重复伤害。[①] 相称性原则作为交战正义的重要体现，要求在攻击之前，用预期从行动中获得的军事优势来衡量对平民的预期伤害[②]，其本质是追求道德成本和道德收益的相称性，即根据战争目标和实际情况使用相称的暴力手段，避免造成过度的“附带伤害”。

在技术乐观主义者眼中，自主武器系统堪称满足相称性原则的典范。一是空间相称，依托人工智能技术实现目标自主锁定，精确限定作战区域，选择性杀伤目标，最大限度避免“附带伤害”；二是时间相称，以美国 X－47B 空天战斗机为代表的自主武器系统，能够实现“半小时覆盖全球”的快速打击，避免“马拉松”式消耗战；三是伤亡相称，自主武器系统实现人与武器的实质性分离，且按照智能作战推演优化作战方案，使实际伤亡无限趋近于零。因此，美国智库兰德公司宣称，自主武器系统比传统军事力量能更好地遵循相称性原则。[③]

相对于美国军方的乐观主义态度，伦理学界以理性的态度分析问题。对于军事行动是否遵循相称性原则的道德评估，在很大程度上取决于参战人员对生命价值和具体情势的主观估计，本身是一个高度复杂、难以进行简单模仿的思维过程。有学者认为，人工智能系统无法理解人类生命的价值，因为它们缺乏拥有生命或感知自己

① 周桂银. 国际政治中的外交、战争与伦理［M］. 南京：南京大学出版社. 2018：287.

② UMBRELLO S, TORRES P. The future of war: could lethal autonomous weapons make conflict more ethical? ［J］. AI & Society, 2020, 35 (1): 273－282.

③ DAVIS L E, MCNERNEY M J, CHOW J S, et al. Armed and Dangerous? UAVs and U. S. Security ［R］. RAND, 2014.

死亡的经验。[①] 如果期望人工智能系统产生类似于人类相称性评估的道德结果，则必须对道德规则进行定量化编码并将其引入训练数据集中，从而在创建的人工智能系统中复制人的道德推理和横向思维。[②] 就目前人工智能技术的发展水平而言，这无疑是难以实现的。当前，西方军事强国利用人工智能技术进行的相称性评估，实质上是以数据层面的相称性替代道德价值层面的相称性，甚至将击毙敌人的数量与战争的正义性等量齐观。这样片面数据化地执行相称性原则，实质上是将复杂的道德推理简单化，其结果必然会造成人道主义灾难的泛滥。

自主武器系统虽然在数据处理与计算速度方面拥有人类难以比拟的优势，但在进行相称性的道德评估方面仍存在诸多瓶颈。沙瑞尔从简单环境与复杂环境的比较视角指出，自主武器系统仅能评估太空、深海、极地等简单环境下的相称性问题，却难以在复杂环境下遵循相称性原则。他指出："在人口稠密的地区，要使自主武器符合'相称性原则'就变得十分困难。如果攻击有效目标造成的附带伤害与该目标的军事价值不相称，那么根据国际法，这种攻击行为应当被禁止。……最难实现的是让机器自身做出决定，判断攻击所需要的比例。这将要求机器扫描目标周围的平民区域，估计潜在的附带伤害，然后判断是否应该进行攻击。这个过程很难实现自动化。"[③] 在复杂环境下，权衡军事目标价值与预期平民伤亡之间的相称性关系，涉及极其复杂的道德推理和评估能力，这是当前自主武

① SKERKER M, PURVES D, JENKINS R. Autonomous weapons systems and the moral equality of combatants [J]. Ethics and Information Technology, 2020, 22: 197 - 209.

② KANUCK S. Humor, ethics, and dignity: being human in the age of Artificial Intelligence [J]. Ethics & International Affairs, 2019, 33 (1): 3.

③ 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：288.

器系统所难以企及的。

钱铖等人从物理伤害与心理伤害的视角指出，军事行动既要在物理毁伤层面遵循相称性原则，也需要在心理伤害层面遵循相称性原则。自主武器系统能够对物理毁伤进行精确评估，进而衡量物理层面的相称性问题，却难以准确评估对敌国民众造成的持久性心理创伤。他们将美国无人机“定点清除”行动给所在国家民众带来的心理层面的创伤称为“雄蜂的阴影”。[①] 在笔者看来，“雄蜂的阴影”反映了当前美国利用自主武器系统反恐面临的道德困境。在某些西方媒体的宣传解读下，无人机、自主武器系统等半自主或自主作战平台以精准的物理摧毁来占据道义的制高点。媒体大肆宣扬依托无人机清除恐怖分子首领的“辉煌战果”，却刻意掩饰阿富汗、也门、索马里、叙利亚等反恐区域内民众反美情绪日益高涨、恐怖主义源头难以清除以及“三股势力”活动愈演愈烈等一系列问题。可以想象，在未来自主武器系统成为执法工具的时代，当局者可能对其快捷高效的执法能力赞誉有加，而普通民众会对“机器人暴政”产生强烈的心理恐惧。

奥特曼等人从战术和战略的视角指出，自主武器系统即便在战术层面能够遵循相称性原则，在战略层面也难以遵循相称性原则。针对少数美国学者提出利用“蜂群”攻击敌对国家的核武器输送系统，以提升美国弹道导弹防御能力的方案，奥特曼等人认为，自主武器系统执行的任务虽然只是小规模战术行动，但是可能造成战略影响。[②] 一旦自主武器系统与核武器攻防技术深度融合，就必然会增

① 钱铖，石斌．“雄蜂”的阴影：美国无人机作战对当代战争伦理的挑战［J］．世界经济与政治，2013（8）：86.

② ALTMANN J，SAUER F. Autonomous weapon systems and strategic stability［J］. Survival，2017，59（5）：130－132.

加核战略对手之间的紧张和不信任，进一步刺激各国优先发展先发制人的自主武器系统，降低核大国之间的战略互信，威胁全球核战略稳定性。

在笔者看来，技术乐观主义者片面强调自主武器系统在简单环境、物理毁伤和战术层面遵循相称性原则的优势，刻意忽视其在复杂环境、心理伤害和战略层面违背相称性原则的问题，本质上反映了对相称性原则内涵的曲解。沃尔泽指出："我们须将'造成的损害'——这想必不仅包括对个人的直接伤害也包括对人类长期利益的损害——与损害对胜利目标产生的促进作用加以权衡比较。"① 在他看来，相称性原则在时间维度包括两方面含义：一方面是短期相称性，另一方面是长远相称性。自主武器系统在简单环境、物理毁伤和战术层面的行为影响都是短期的，而在复杂环境、心理伤害和战略层面的行为影响将是长期持续的。两者相较，维护人类持久和平的长期利益无疑更为重要。

（三）对战后正义的挑战

战后正义又称战争责任正义，是正义战争理论的重要组成部分。战后正义原则主要关注人类应当承担的战争责任，具体体现在战后秩序重建、侵略罪行审判、战争赔偿以及制裁等战争责任分配问题。人类既是享有发动战争权利的主体，又是承担战争责任的主体。正如沃尔泽所指出的："责任分配是对正义论的重要检验。……只要我们准确地指出责任者，或者至少我们的责任分配和道德判断符合战争的实际经验、对战争的所有痛苦保持敏感，就会极大地增强正义

① 沃尔泽. 正义与非正义战争：通过历史实例的道德论证［M］. 任辉献，译. 南京：江苏人民出版社，2008：143.

论的力量。如果最终无人负责，在战争中就没有任何正义可言。”[①] 依据战后正义原则，人类在享有战争权利的同时也必须承担战争责任，所有的战争罪犯都必须追究责任。在智能化战争模式下，伴随自主武器系统的出现与应用，人作为战争的责任主体从作战系统中脱离出来，将生死攸关的决定权交给机器。对于这一变化，多彻蒂、舒尔茨克、马利克等战争伦理学者认为，自主武器系统无法识别、承担责任并接受惩罚，原有的战后正义规约难以约束一场完全由机器人来进行的战争。在一场智能化武器装备主导的战争中，如何确定战争责任主体？如何公正地分配战争责任？军用人工智能对于战后正义的挑战具体体现为以下几个方面。

1. 挑战战争责任主体的唯一性原则

在传统的战争模式下，人类是承担战争责任的唯一主体。随着人工智能的军事化发展步伐不断加快，各类自主武器系统取代人类成为战争行动的决策者和执行者，由此也冲击了人类作为战争唯一责任主体的地位。客观而言，基于自由意志是制定道德标准的先决条件，因此，只有人类才具有遵守战争伦理和国际人道法的能力。然而，具有完全自主模式的武器系统可以在无人干预条件下实现自主感知、决策和行动，并将人类完全剥离系统回路。自主武器系统是否能够具备在非结构化战争场景中实现态势感知并做出适当的道德判断的能力？能否与传统战争责任主体一起，在道义层面共同扮演战争责任主体的角色，进而将未来战争引向更为人道的方向？目前仍然存在巨大的不确定性。

① 沃尔泽．正义与非正义战争：通过历史实例的道德论证［M］．任辉献，译．南京：江苏人民出版社，2008：319.

2. 挑战战争责任分配的匹配性原则

沃尔泽指出："如果有可以辨认的战争罪行，就必定有可以辨认的罪犯。"[①] 在传统作战模式下，人类直接操控武器系统（人在回路中）或者严密监督武器系统的运行（人在回路上），人与武器系统的关联是紧密的。因此，人类操控员或武器研发者需要对武器操控失误引发的人道主义灾难承担不可推卸的责任。然而，对于具有高度自主性的智能化武器装备而言，其在作战过程中无须任何操控者，其研发是由来自全球不同国家、不同领域、不同部门的研发团队在相当长的一段时间内共同合作完成的，因此很难甚至不可能将战争责任与战争主体一一对应起来。如果一个自主武器系统在作战过程中出现了意料之外的运行错误，并随之引发严重的人道主义灾难，战争责任分配必将陷入困境，无论是自主武器系统的设计者还是其指挥官，都不能独立地对其造成的战争后果承担道德责任。确定哪些责任应当由其设计和测试者承担？哪些责任应当由自主武器系统承担？确定谁对失误负责或者应该由谁负更大的责任？上述问题目前仍然存在巨大的争议。

人与武器分离还会引发"问责空白"风险。在传统战争模式下，人类作为军事行动的遂行者和武器的直接操控者，其所从事的战争行为与其导致的后果之间存在清晰的对应关系。正如沃尔泽所说："如果有可以辨认的战争罪行，就必定有可以辨认的罪犯。"随着自主武器系统的军事应用，人类在战争系统中的定位由"人在回路中"逐渐向"人在回路上"乃至"人在回路外"转变，作战主体与作战手段之间的距离日益拉大。当前西方军事强国广泛运用的"捕食者"

① 沃尔泽. 正义与非正义战争：通过历史实例的道德论证［M］. 任辉献，译. 南京：江苏人民出版社，2008：319.

“收割者”等半自主武装无人机，就是典型的“平台无人、体系有人”作战模式。无人机操控员在千里之外实现对目标的远程打击，他们感官上与战场隔绝可能导致其行为与责任的分离，作战人员通过计算机按键与远程视频反馈来实施打击，因此存在以游戏心态来杀人的风险。可以预见，一旦真正意义上的自主武器系统大规模投入战场，人类将从瞄准和击杀的作战链条中被移出，将从根本上改变人与战争、人与杀戮之间的关系，也将彻底打破正义战争理论所强调的“战争罪犯”与“战争罪行”之间的对应关系。美国亚利桑那州立大学罗夫和卡内基梅隆大学丹克斯指出：“人类对于自主武器系统的信任源自人类对其行动可预测性的判断。”① 如果自主武器系统错误的杀戮行为超出了人类可以预测的范围，就可能产生人类不需要对杀戮行为负责任的结果，这种“问责空白”将可能导致更多的杀戮和痛苦。

自主武器系统的作战应用，不仅会造成机器难以代替人类承担战争责任的“问责空白”，而且可能导致战争责任的分配困境。在传统战争模式下，责任主体构成相对简单，主要由政治家、军队指挥官和士兵等特定对象构成。自主武器系统研发链条的复杂性和高技术密集性，促使程序员、算法工程师、数据工程师及自主武器装备生产商等新兴责任主体出现。自主武器系统研发过程的复杂性与作战应用过程的复杂性双重叠加，导致战争责任分配面临巨大的困境，甚至可能出现“责任转嫁”的风险。一方面，使用自主武器系统的国家可以通过匿名攻击的方式，使得自主武器攻击难以溯源，进而逃避战争罪责。例如，2010 年发生的“震网”病毒攻击伊朗核电站

① ROFF H M, DANKS D. “Trust but verify”: the difficulty of trusting autonomous weapons systems [J]. Journal of Military Ethics, 2018, 17 (1): 1.

事件、2020 年发生的伊朗核科学家遇刺事件，都是采取自主武器系统匿名攻击的方式，使得攻击者可以逃避战争责任的追究。另一方面，被攻击国家可以利用智能化作战平台对作战网络的高度依赖，采取电磁干扰、植入木马、黑客入侵等方式，操控其进行违背人道主义准则的军事行动，再将罪责转嫁给使用智能化作战平台的国家。2011 年，伊朗电子战专家成功俘获美军 RQ－170“哨兵”无人机的案例，说明在自主武器领域的技术先发国家与后进国家一样，都可能面临战争“责任转嫁”的风险。

3. 挑战战争权利与责任的对应性原则

从权责匹配的角度来考量，社会对于军人职业身份的尊重和认可源自其所承受的责任，军人必须以自我牺牲精神承担保护民众的责任，这与军人在社会资源分配中所享有的优先权利是相互匹配的。当前，学界对于智能化战争的战后正义的研究，主要从伦理规约的层面，探讨人工智能技术的军事应用对战争责任分配带来的挑战与应对。然而，在笔者看来，实现战后正义，不仅要靠法律法规和伦理原则的约束，而且需要将承担战争责任作为一种文化进行传承。正如麦克阿瑟所述：“军人必须承担保护弱者和没有武装者的责任。这是军人担任这个神圣职务的根本理由。”① 军人作为一种社会尊崇的职业存在的道德基础，就在于军人以自我牺牲的精神替代平民承担更大的责任与风险。军人具有“保家卫国”“不怕牺牲”“英勇杀敌”等武德精神，正是责任文化传承的鲜明体现。一旦自主武器系统取代人类成为战争主体，战争将会进入“后英雄主义时代”，各类自主武器系统将承担绝大部分的作战风险，战争中的军人被置于比

① 沃尔泽. 论战争［M］. 任辉献，段鸣玉，译. 南京：江苏人民出版社，2011：33－34.

平民还要安全的境地。情况正如美国空军飞行员里扎所指出的："智能化战争缺失了对军人个人意志的考验，我们在战争中找不到任何意义。"① 这种安危关系的反转，必将影响形成军人身份认同的责任价值基础，甚至危及整个社会对于军人职业的道德评价，即他们将不再被期待具备过人的勇气和大无畏的精神，从而在无形中削弱了确认军人职业身份的道义根基，也意味着对军人武德精神的挑战。

五、应对安全风险的启示

习近平总书记在十九大报告中指出："加快军事智能化发展，提高基于网络信息体系的联合作战能力、全域作战能力，有效塑造态势、管控危机、遏制战争、打赢战争。"习近平总书记在二十大报告中进一步指出，"增强新域新质作战力量比重，加快无人智能作战力量发展"。在全球无人化、智能化战争方兴未艾的背景下，作为军用人工智能领域的追赶者，中国既要努力学习世界军事强国发展军用人工智能的成功经验，做好打赢未来智能化战争的充分准备，同时，也要清醒认识到人工智能军事应用潜在的危机与挑战，积极应对全球人工智能领域军备竞赛可能带来的国际安全风险。

（一）学习和借鉴军事强国推进人工智能发展的成功经验

一方面，要学习战略规划先行式发展思路。美国军用人工智能发展的一个鲜明特色，就是高度注重战略规划和顶层设计，通过构建"国家—军队—军种"三位一体、层层对接的战略规划体系，组

① RIZA S. Two - dimensional warfare: combatants, warriors, and our post - predator collective experience [J]. Journal of Military Ethics, 2014, 13 (3): 257.

建专职化军事智能化战略管理机构，牵引武器装备、编制体制、作战训练、军事理论、人才培养等方面工作的整体推进。这启示我们要未雨绸缪地强化发展军用人工智能的战略设计，既要对标世界一流军队，对照打赢智能化战争要求，又要结合国情和军情，因地制宜、科学合理地制定战略规划。我们应当充分认识到我国高度集中统一的人工智能举国发展模式所具有的制度优势和效率优势，充分认识到我国民用智能化应用领域的高速发展态势和巨大发展潜力，构建军民协同创新的人工智能战略发展格局，高效整合军地资源，真正实现多位一体、协同推进、跨越发展，用人工智能牵引我军机械化、信息化向更高水平、更深层次加速发展。另一方面，要实现军用人工智能发展在行动域和认知域的齐头并进。美国在军用人工智能领域的全球领导地位，既体现为行动域内智能化武器装备等作战手段居于世界领先水平，也体现为认知域内谋求作战理念创新和制定作战规约的全球领导权。这也启示我们既要下大力气推动智能化“撒手锏”武器的研制与应用，也要把握人工智能对作战形态、作战方式发展变化的影响，着力加强智能化无人作战、人机协同作战以及智能化“蜂群”作战等作战方式、方法研究，积极参与国际人工智能军控会议的交流、研讨与规则制定，努力实现人工智能在行动域与认知域的双向互动发展。

（二）准确认识军事强国推进人工智能发展的优势与挑战

美国在国防领域长期秉承技术制胜的战略传统，为人工智能的军事化应用培育了文化土壤。作为世界第一军事强国，美国自第二次世界大战以来，始终将技术制胜作为其军事战略设计的显著标识符。长期以来，美军高度重视前瞻性、先导性、颠覆性科学技术的军事应用，将其作为美军生成新质战斗力的“倍增器”。正如曾任美

军参谋长联席会议副主席的迈亚将军所指出的："技术领先是美国武装力量的一个标志。"从"二战"期间原子弹的研制，到冷战期间阿帕网的诞生，再到海湾战争期间 F－117 隐形战斗机、"战斧"精确制导弹药和全球卫星导航定位系统等信息化武器装备的综合应用，技术制胜的文化传统一直是牵引美军军事创新最持久而弥坚的源动力。当前，美国推出"第三次抵消战略"，加速人工智能的军事化应用进程，同样是技术制胜传统的延续。美国《2018 年度国防部人工智能战略概要》明确指出："美国军方最大的优势之一就是我们部队的创新性，最具变革性的人工智能功能将来自军事领域的'前沿'实验。"美军对于科技前沿的高度敏锐感、对于追逐颠覆性武器装备的强大热情以及不计工本的军事科研传统，无疑为加速军事智能化的进程起到潜移默化的积极影响。

美国维护世界霸权的现实需求，为其推进军事智能化应用进程提供了强劲动力。美国是世界第一军事大国，不仅拥有全球部署的庞大军事力量，而且承担着包括常规战争、特种战争、有限战争和"反恐战争"在内的各种不同类型的任务。21 世纪以来，美国为维护国家利益和巩固全球霸权，先后参与了阿富汗战争、伊拉克战争、利比亚战争和叙利亚的武装冲突，并在全球范围内开展"定点清除"行动，多样化的军事任务也为人工智能领域的军事应用提供了需求牵引。例如，在"反恐战争"中，美军作战部门所面临的一个关键问题，即如何在一个高度机动、混杂、变幻的战场环境中保证对恐怖分子的"发现即摧毁"，由此催生了"察打一体"无人化作战平台的出现，"捕食者""收割者"等武装无人机迅速发展成为美军的反恐利器，也为美军作战模式由"人在平台之中"的传统信息战模式向"平台无人、系统有人"的智能化远程遥控作战模式转型提供了契机。当前，人工智能技术已经广泛应用于美军情报搜集、监视

和侦察，后勤保障，网络空间行动，信息操纵、指挥和控制，半自主式和自主式运载工具，以及致命性自主武器系统等多个领域，并在其遍及全球的军事力量和作战任务中得到应用、检验、完善和普及推广。可以肯定的是，美国的军事智能化将在这一过程的持续循环中接受检验并且持续快速发展。

同时，美国在实施军用人工智能战略方面也面临以下挑战。

一是国际竞争的挑战。人工智能是引领未来战争的颠覆性技术，世界军事强国纷纷将发展军事智能化作为维护国家安全、迎接新军事变革和打赢未来战争的重大战略举措。随着全球军用人工智能领域军备竞赛日趋白热化，美国在军用人工智能领域面临的发展阻力也逐渐增加。一方面，由于全球主要军事强国对于军事智能化发展的重视程度与日俱增，国际上围绕军用人工智能研究、开发和人力资源的争夺也更加激烈，美国军方在保留和招揽全球军事智能化高端人才方面受到挑战，这将直接影响美国在军事智能化领域的持续竞争力。另一方面，国际竞争也在制度规则层面对美国军事智能化发展造成掣肘。如何制定军用人工智能研发、应用的国际规则，这是国际上围绕军用智能化领域竞争的焦点之一。与美国倡导加速推进人工智能在国防各领域的广泛应用截然不同，欧盟和部分技术后发国家主张对军用人工智能技术进行严格的法律规制。2018 年 9 月，欧洲议会通过禁止使用“杀手机器人”的决议，[①] 对美国试图领导军用人工智能国际规则的制定构成了严峻挑战。美国如何在全球军事智能化的国际竞争中突围，谋取军事智能化领导者地位，仍存在较大变数。

① CONN A. European parliament passes resolution supporting a ban on killer robots [EB/OL]. (2018 - 09 - 14) [2022 - 04 - 07]. https://futureoflife.org/2018/09/14/european-parliament-passes-resolution-supporting-a-ban-on-killer-robots/.

二是军民技术转移的障碍。美国将民用人工智能领域的研发优势转移到军用领域，面临技术标准、组织体系和适用范围等方面的障碍。首先，美国在军民技术领域的具体标准方面存在一定的差异。较之民用领域，军用领域对于技术性能和安全性的要求更高，对于数据权利和知识产权方面亦有严苛、烦琐的审批流程，从而在无形中影响了民用人工智能技术向军事领域转移的进程。其次，在组织体系层面存在阻力。军事智能化可能导致军队组织构成的颠覆性革命。美国国会研究处指出：如果美国军方只是简单地使用人工智能来逐步改进现有的流程或加强当前的作战观念，那么其所面临的阻力并不大；然而，如果军方尝试释放人工智能的潜力来改变战争面貌和作战样式，就必须改变现有军队的基础性组织架构，可能导致美军内部强烈抵制人工智能的应用。[①] 最后，在民用环境中开发的人工智能系统转移到军用环境，也可能遭遇适用范围的障碍。军用环境系统的组织构成、行为手段、影响因素、演化进程较之民用环境具有其特殊性和复杂性，智能化系统在两种截然不同的运行环境之间切换需要进行重大调整。

三是军事应用的潜在风险。美国推进军事智能化面临多方面的风险。首先是技术风险。美国国会全球安全分析师塞尔认为，军用人工智能系统可能为敌对国家提供技术漏洞和网络入侵途径。由于军用人工智能系统的运行主要基于应用软件，因此更为容易遭到计算机黑客的信息窃取和网络攻击。随着各类人工智能驱动的自主武器系统被广泛应用于军事领域，美军作战网络面临对手“瘫痪”威胁的可能性也不断增长。其次是伦理、法律风险。美国《2018 年度

① Congressional Research Service. Artificial Intelligence and National Security [R]. Congress, 2019: 27.

国防部人工智能战略概要》指出，部分民用公司基于伦理、法律等方面考虑拒绝与军方就人工智能展开合作。对硅谷几家顶级人工智能公司领导层的调查问卷显示，近 80% 的参与者认为军方与民用技术企业之间存在诸多不信任，其中一个重要的因素就是基于伦理、法律风险的考量。[①] 例如，由于军用人工智能项目可能对个人隐私权构成侵犯，谷歌公司退出美国国防部 Maven 项目计划，并禁止其收购的人工智能企业参与军方的合作项目。最后是舆论风险。目前，围绕人工智能技术特别是致命性自主武器应用于战争的问题，美国国内和国际舆论都存在较大争议，联合国、红十字国际委员会等国际组织，以及“国际人工智能联合会议”“杀手机器人禁令运动”“未来生活研究所”等非政府组织均旗帜鲜明地反对“杀手机器人”的研制、开发，从而在国际舆论层面对美国军事智能化进程造成强大的压力。

（三）积极应对人工智能军备竞赛可能带来的国际安全挑战

当前，人工智能引发的军备控制、危机管控，以及伦理、法律风险都是全球性问题，需要世界各国共同努力来应对和解决。一方面，中国作为联合国安理会常任理事国，应当深度参与并积极推动联合国框架下围绕人工智能和自主武器系统的军备控制进程，努力将致命性自主武器系统的军控议题以专题单元方式列入联合国《特定常规武器公约》会议，积极呼吁和推动致命性自主武器多边军控谈判级别由当前的政府专家组会晤上升至主权国家政府间谈判，以促使相关军备控制国际公约的签订。另一方面，围绕人工智能技术

① Congressional Research Service. Artificial Intelligence and National Security [R]. Congress, 2019: 18.

作战应用可能导致的危机管控和意外战争风险，我们应当充分利用组织和参与香山论坛、香格里拉峰会等国际对话机制的平台优势，向国际社会阐明禁止研发和应用致命性自主武器系统的鲜明立场。同时，以磋商、谈判和领导人热线交流等多种形式，在国际社会就人工智能可能引发的全球战略稳定性挑战形成战略共识，探索搭建起迅捷、高效的战略沟通与危机管控机制，努力确保将人工智能技术的作战应用置于各国可以控制的范围之内，杜绝意外战争或危机无限制升级的潜在风险。此外，我们还应当汇聚全球热爱和平的科学家、非政府组织和民间团体的合力，共同针对人工智能军事应用产生的伦理、法律问题构建全球治理网络，组织科学家、伦理学家和法学家通过主流媒体门户网站和社交网络等传播渠道积极发声，开展关于人工智能军事应用风险的科学、伦理和法律知识普及，发起全面禁止研发和应用“杀手机器人”的全球性倡议，赢取民众对于人工智能军事应用进行全球治理的广泛支持，以强有力的国际舆论压力敦促各国军用人工智能发展符合战争伦理和战争法的规约。

第三章
信息搜集与网络情报战

所谓情报战，指的是特定国家、特定利益集团受到政治、经济或军事利益驱动，采取多样化手段方式，有目的、有意识地窃取和利用敌方的情报信息，同时对敌方的情报搜集行动加以防范，达成获取敌方核心关键信息、削弱或摧毁敌方情报信息搜集利用能力等目的。情报战也是古往今来人类战争中一种常用的作战方式。在智能化战争中，围绕情报信息的攻防作战已经成为重要的作战样式。无人机反恐作战并非单一的火力打击行动，而是涵盖情报搜集、分析决策和火力打击三个主要环节的综合系统。其中，情报搜集是反恐作战的初始环节。本章重点关注因使用智能化手段进行反恐情报搜集而带来的伦理问题。

网络情报战的兴起对于反恐作战伦理研究的影响，既体现为它是军事技术冲击现代战争伦理的典型案例，又体现为它预示着人类

战争模式由 War 1.0 向 War 2.0 的嬗变，由此可能引发整个战争伦理体系的革命性变化。

较之传统作战方式，当前方兴未艾的网络情报战没有时空限制，隐蔽性强，攻击范围广且效费比高，其搜集的情报对于提升智能化战争的作战效率无疑产生了积极的影响。然而，网络情报战所导致的一系列伦理与法理问题引发了学界的激烈争论。作为一种颠覆性的作战手段，网络情报战并不直接杀伤作战人员，这是否构成其符合人道主义原则的凭据？鉴于越来越多的国家将网络空间反恐作战纳入国家战略层面，网络军事化已经成为不争的事实。讨论网络情报战的正义性问题，不仅具有理论层面的价值，而且具有现实的政治意义。

国内学界朱雁新[①]、刘正[②]从国际法的视角对网络战的正义性进行了探析，对于网络反恐情报战与战争伦理的关系尚缺乏有深度的专题研究。国外学界对网络战的讨论涉及法理、伦理、政策、技术和战略等多个维度，然而，囿于研究主题、研究视阈与价值取向的差异性，加之因网络情报战手段创新不断引发新的伦理问题，当前的研究尚不深入和系统，尤其是网络攻防技术对美国反恐作战方式和战争伦理原则的深层次影响，关注力度不够。

本章回顾了网络情报战的发展历程，在此基础上，从经典正义战争理论的分析框架出发，分别讨论当前美国网络反恐情报战对开战正义、交战正义及战后正义相关伦理原则所产生的冲击与挑战。

① 朱雁新．计算机网络攻击之国际法问题研究［D］．北京：中国政法大学，2011.

② 刘正．网络战的国际法应对［J］．山东社会科学，2006（3）：139－141.

一、信息搜集和网络情报战的途径方式

面对遍布全球的情报侦察系统，恐怖组织的武装分子通常伪装成平民且行踪诡异，传统空间的情报搜集面临严峻挑战。情报部门纷纷将目光投向新兴的网络空间，特别是社交网络。随着恐怖主义势力向网络空间特别是各类社交媒体的广泛渗透，反恐作战的作战环境也相应地由实体空间向虚拟的网络空间拓展。“9·11”事件之后，美国构建了以“棱镜”计划为代表的智能化情报搜集系统，在网络空间进行情报搜集。依托人工智能的网络反恐情报战日益发挥着至关重要的作用。具体而言，情报人员依托网络攻防技术手段，利用大数据、人工智能技术识别恐怖分子身份，获取恐怖主义袭击活动的情报信息，对恐怖组织及“恐怖主义国家”的网络渗透实施信息压制、瘫痪情报信息系统和攻击关键基础设施等，从而有效配合在实体空间进行的反恐作战。

在战争领域利用网络空间的历史可追溯至1991年的海湾战争，美军以植入染毒芯片的方式，致使伊拉克的防空系统陷入瘫痪。进入21世纪，一方面，由于互联网技术的不断发展及国际社会对其依赖度持续增长，网络空间作为第五维空间的战略意义日益凸显，利用人工智能技术进行网络情报搜集与争夺制网权，成为美、俄等军事强国的共识；另一方面，恐怖组织加快向网络空间渗透的步伐，在此背景下，各国军方和安全部门运用各种人工智能技术手段遂行网络情报战成为大势所趋。

近年来，各国政府和军方大力加强基于社交媒体账号的情报搜集方式方法创新。美国空军的一位情报官员指出：“我们的搜集工作

必须利用包括社交媒体在内的各类网络工具，以跟上敌人的步伐。”[①] 美国政府鼓励支持谷歌和雅虎搜索引擎的新的本土智能情报搜集平台，通过扫描其互联网浏览历史来识别潜在的恐怖分子。这些监视中心及其服务器应在中央情报局或武装部队的严密监视下进行。谷歌已经开发出了重定向方法，使用搜索算法、关键字和有暴力倾向的人经常使用的短语。[②] 2017 年，脸谱、微软、推特等公司宣布成立全球反恐论坛，为对恐怖分子和暴力极端主义分子怀有敌意的群体提供情报服务。[③] 借助这一论坛，各国政府关注恐怖分子社交网络中异常活跃的社交账号，梳理其关注与联系对象，搜集恐怖分子在社交账号上聊天、交友、购物，以及上传视频、音频、图片的信息，为锁定和追捕恐怖分子提供情报支持。Voyager Labs 指出，强大的人工智能能力不仅体现在使用开源互联网络方面，而且体现在可以非常有效地分析“暗网”帖子和通信方面。一些人工智能程序在识别网上购买模式方面非常成功，其参数包括物品类型和数量，这些参数通常与恐怖活动相关。物品通常包括武器、弹药、武器配件、可用于炸药的化学品，甚至与人口贩运相关的商品（有时用于资助恐怖活动）。美军已经在反恐作战中成功应用网络情报战的成果。2015 年 6 月 8 日，美国空军空战司令部司令卡莱尔将军证实，美军情报部门依据一名“伊斯兰国”武装分子在推特上的自拍照，

① HOFFMAN M. US air force targets and destroys ISIS HQ building using social media [EB/OL]. (2015 - 06 - 03) [2022 - 04 - 07]. https://www.military.com/defensetech/2015/06/03/us - air - force - targets - and - destroys - isis - hq - building - using - social - media.

② KAUSHIK N. Artificial Intelligence in counter - terrorism [EB/OL]. (2021 - 10 - 26) [2022 - 04 - 07] http://www.youngbites.com/newsdet.aspx?q=362464.

③ Xinhua. US social media networks, Microsoft join hands to fight against terrorism [EB/OL]. (2017 - 06 - 27) [2022 - 04 - 07] https://usa.chinadaily.com.cn/world/2017 - 06/27/content_29901099.htm.

成功锁定并用无人机发射导弹摧毁了该组织的一个总部大楼。[①] 这成为依托网络情报搜集支撑无人机反恐作战的一个典型案例。这个案例体现出近年来实行网络空间情报作战的基本途径与方式。

各国持续推动社交媒体平台的反恐舆论战与网络信息压制。针对“伊斯兰国”等恐怖组织在社交媒体开展的极端主义宣传，采取针锋相对的反制措施。例如，印度在军队和民间情报机构中建立了负责协调的人工智能情报室，以期利用人工智能方法解决新的安全问题。印度陆军人工智能情报室提高其在国内的可操作性，并与盟国（如以色列、美国和日本）的人工智能部门合作。通过军队的定期媒体简报会，提高社会对政府安全措施的接受度。为了阻止恐怖袭击，还鼓励公民在社交媒体上进行监视和提高警惕。[②] 美军推动“在线虚拟身份管理”项目的经验，发展社交媒体平台的舆论战力量，开发多语种舆论对抗倍增软件，在强化己方舆论输出的同时，对暴力恐怖信息与极端思想进行高效过滤。针对“圣战”宣传对社交媒体用户的认知、情感、意志等心理过程的负面影响，加强对相关情绪启动效应的研究，着重研究极端思想对“独狼”式袭击者的心理催化过程，采取隐蔽的阈下心理反制措施，通过在恐怖组织的宣传内容中植入不易察觉的阈下信息，对其负面情绪的传递实施反向干扰。

值得注意的是，近年来，网络反恐情报战的手段方式不断创新，日益突破“软杀伤”的限制，开始研发、使用的新型网络武器具有

① HOFFMAN M. US Air Force Targets and Destroys ISIS HQ Building Using Social Media [EB/OL]. (2015-06-03) [2022-04-07]. https://www.military.com/defensetech/2015/06/03/us-air-force-targets-and-destroys-isis-hq-building-using-social-media.

② KAUSHIK N. Artificial Intelligence in counter-terrorism [EB/OL]. (2021-10-26) [2022-04-07]. http://www.youngbites.com/newsdet.aspx?q=362464.

实体破坏能力，既能够搜集反恐情报，又能够瘫痪对方信息网络设备和国家关键基础设施。伊朗一直是美国构建反恐情报网络的重点覆盖区域。据《纽约时报》报道，“震网”病毒是以色列和美国合作进行的“奥林匹克”计划的一部分，其首要目的在于搜集伊朗利用核技术实施恐怖主义袭击的情报，并通过破坏关键核设备迟滞伊朗核恐怖主义计划的进程。“震网”病毒对伊朗布什尔核电站内部控制系统的攻击，成为网络反恐应用的里程碑事件。在此之前，各国政府和军方在网络攻击的对象设置上基本局限于计算机系统和互联网络，“震网”病毒则开创了网络武器直接攻击物理世界基础设施的先河。“震网”病毒利用西门子 WinCC 系统和 Windows 系统的多个漏洞遂行攻击指令，伪装 Realtek 与 JMicron 两大公司的数字签名，顺利绕过安全检测，自动锁定并攻击工业控制系统软件，以控制设施涡轮机或冷却系统运作，在工作人员毫无察觉的情况下让设备失控自毁。俄罗斯卡巴斯基实验室对“震网”病毒进行反编译后指出：“震网”结构非常复杂，不可能是纯粹的黑客行为，除非有国家和政府的支持和协助，否则很难发动如此规模的攻击。[①] 攻击伊朗核设施的“震网”病毒，与攻击伊朗工业控制系统数据的“毒区”病毒，以及攻击伊朗石油部门的商业情报的“火焰”病毒，共同构成了美国网络反恐武器的“攻击集群”。

二、网络情报战与开战正义

迄今为止，网络情报战作为一种新的战争样式，被应用于不同

① Kaspersky. Kaspersky Lab provides its insights on Stuxnet worm [EB/OL]. (2010-9-24) [2022-04-07]. https://www.kaspersky.com/about/press-releases/2010_kaspersky-lab-provides-its-insights-on-stuxnet-worm.

的两类战争之中：一类是在陆、海、空等自然维度展开的常规战争，如俄罗斯发动的针对格鲁吉亚的俄格战争，以及美国发动的阿富汗战争、伊拉克战争，网络情报战扮演辅助作战的次要角色；另一类是在网络维度展开的战争，即惯常所说的“网络战争”，网络攻防技术是唯一的作战手段。就前者而言，网络反恐情报战对于宣布开战的伦理考量并无直接影响，然而就后者而言，网络攻防无疑成为一种独立的作战样式。网络情报战对开战正义的冲击，主要体现在正当理由、合法权威和最后手段三个方面。下面主要分析前两个方面。

在正义战争理论中，发动战争的唯一正当理由，即是对他国的侵略行为进行自卫。网络武器作为情报战手段出现，对正当理由原则构成了冲击。一国在他国网络空间搜集情报信息的行为，是否属于侵略行为？依据 1974 年联合国大会通过的《关于侵略定义的决议》：“侵略是指一个国家使用武力侵犯另一个国家的主权、领土完整或政治独立，或以本定义所显示的与《联合国宪章》不符的任何其他方式使用武力。”① 由此可知，问题的症结在于网络空间究竟是归属各国的主权范围，还是世界共享的全球公域。从历史发展的视角来看，互联网创建的宗旨是开放互通与资料共享，实现全人类信息财富的共同管理与利用，始终是推动互联网发展的源动力，因此，互联网无疑具有全球公域的特征。然而，从国家安全的视角来看，信息安全已成为国家安全的重要方面，提高网络空间的安全防护成为信息时代维护国家安全的必然选择。可以预见，随着国际政治从领土、领空扩展到网络空间，国家主权从领土、领空扩展到“信息边疆”，围绕网络主权与全球公域的争议仍将延续。

网络情报战是否符合正当理由原则的问题，即何种类型的网络

① 王铁崖. 国际法［M］. 北京：法律出版社，1981：73.

情报搜集方式，构成了对他国主权的侵略行为？这一问题可细分为三类情况。一是在网络空间搜集情报的行为已实施，并且造成了危害。从毁伤机理与杀伤方式来看，网络情报搜集的特殊性在于其并不直接造成人员伤亡。长久以来，人们将对生命的剥夺看作战争暴力本质的集中体现，也是战争与人道主义原则的根本冲突所在。作为不危及人类生命的作战手段，依托网络攻防技术实施情报搜集尚未构成战争行为。然而，有学者针锋相对地指出，网络攻击方式也可以具有攻击性，一旦作战目标是瘫痪对方的情报信息网络，它会造成严重的社会危害。例如，俄罗斯对格鲁吉亚关键信息网络系统实施“拒绝服务式”网络攻击，可能会造成大面积的社会恐慌，进而导致社会危机的爆发。[①] 常规武器对人员的物理杀伤与关键信息系统瘫痪对人类社会的威胁，两者孰轻孰重很难进行评估。二是网络情报战已经实施但尚未造成危害。例如，用于搜集反恐情报的木马程序和蠕虫代码，已经植入敌国政府或军方网络系统，但尚未运行。依据惯常的逻辑，在他国领土范围内埋设地雷是明确的战争行为。以此类推，在他国网络信息空间安装恶意软件是否属于战争行为？对此，2009 年 3 月，俄罗斯国家安全委员会副秘书索科洛夫在表明俄政府关于网络情报战的基本立场时明确指出，禁止敌对国家在计算机系统中嵌入可在日后危急时刻启动的恶意代码。然而，考虑网络反恐情报战的归因困难，如何明确网络病毒植入行为背后的国家意图本身存在巨大的难度。三是具备大规模网络情报搜集能力，但尚未针对他国网络空间或关键基础设施实施网络攻击行为。近年来，部分西方学者将“自卫”概念的解释扩大，提出“预防性战争”的

① LIN P, ALLHOFF F, ROWE N C. War 2.0 cyberweapons and ethics [J]. Communications of the Acm, 2012 (3): 25.

概念，认为先发制人的进攻同样是“自卫”的体现，其目的在于规避后来在环境恶化的条件下进行战争的风险。[①] 2003 年，美国以伊拉克拥有大规模杀伤性武器并准备实施核恐怖主义袭击为由发动伊拉克战争，被美国学者广泛称为“预防性战争”。[②] 从近年来美国网络信息安全战略的演变来看，由克林顿政府围绕基础设施的“全面防御”到布什时代的“网络反恐，攻防结合”再到奥巴马政府的“攻击为主，网络威慑”，与美国网络安全战略“由守到攻”的演变相对应。美国伦理学界也有学者提出将先发制人模式的“自卫”权拓展至网络情报战领域。[③] 其实质是为美国在网络空间展开或即将展开的情报搜集和网络攻击行动寻求道义支持，体现了西方战争伦理功利主义的基本价值取向和现实主义的主要行为准则。

对于开战正义而言，正当理由是必要条件而非充分条件，除了正当理由，还需要符合合法权威原则。探讨网络情报战的合法权威时遇到的一个核心问题，即国际合法权威与国内合法权威的冲突。一方面，现代国际法和《联合国宪章》是以限制国家发动战争和威胁使用武力的权力为主旨。鉴于网络空间的全球公域特性，中国、俄罗斯等国一直倡议在联合国的框架下开展网络安全合作，形成约束使用网络武器和发动网络战争的国际公约，中、俄等国曾于 2011 年 9 月向第六十六届联合国大会提交“信息安全国际行为准则”，引起国际社会广泛关注，推动了信息和网络空间国际规则制定进程。另一方面，少数国家凭借其在信息技术领域的领先地位，将网络情

① MILLIES S, HOWER S. Is a preventative war a just war? [C]. Midwestern Political Science Association, 2004: 1 - 10.

② DIPERT R R. Preventive war and the epistemological dimension of the morality of war [J]. Journal of Military Ethics, 2006 (5): 32 - 54.

③ DIPERT R R. The ethics of cyberwarfare [J]. Journal of Military Ethics, 2006 (4): 392 - 394.

报战作为一种新的战争手段，为其所推行强权政治和霸权主义服务。在伊拉克战争中，美国将伊拉克定义为“恐怖主义国家”，对伊拉克实施情报信息网络瘫痪，利用对互联网名称与数字地址分配机构的控制权，取消了“.iq”的域名，将伊拉克从网络空间中抹除。近年来，美国先后颁布《网络空间国际战略》和《网络空间行动战略》，把自身塑造成网络空间“合法权威”的代言人。在网络空间，掌握网络核心技术的强权国家与技术后发国家之间围绕“合法权威”展开的力量博弈，成为影响网络空间战略稳定性的关键因素。忽视国际法对网络反恐情报战的约束作用，只会加剧网络空间的分歧、冲突与军备竞赛。

三、网络情报战与交战正义

交战正义具体是指战争进行过程中参战各方必须遵循的伦理规则。其中，区分原则和相称性原则构成了战争行为伦理的核心内容。网络情报战作为近年来兴起的一种新的作战样式，其运用过程也需要相应的道德规范来约束。然而，网络情报战对区分原则和相称性原则构成了一定的挑战和冲击，网络情报战在交战过程中是否合乎道义的问题，引发了国际社会的广泛争议。

区分原则关注的是作战中哪些属于合法的攻击目标。在传统的交战正义标准下，区分原则的应用主要集中在居于次要地位却是基础性的“非战斗人员豁免权”原则上，主要涉及非作战人员与作战人员的区分问题。然而，对于网络反恐情报战而言，讨论这一新型作战方式在作战应用中是否遵循区分原则时，首先会遇到攻击目标偏转的问题。较之常规战争，网络情报战的攻击目标不是人，而是电子计算机、信息系统的敏感数据信息，以及国家信息基础设施。

例如，为了搜集情报信息而在网络空间植入木马，不会直接造成人员伤亡，因此表面上并不与区分原则产生直接冲突。基于网络攻防手段发动的网络情报战，似乎可以一劳永逸地规避了困扰现代战争的区分困境问题。然而，事实上，从网络情报战的主要攻击目标来看，它同样无法规避军事设施与民用设施之间的区分困境。一方面，计算机系统和互联网络作为现代信息系统的基本构成要素，本身即具有军民两用的特性；另一方面，某些基于民用目的创建的关键基础设施，同样可能具备转为军事用途的潜能，遭受“震网”病毒攻击的伊朗布什尔核电站内部网络系统即是典型代表。

网络情报战对于区分原则的挑战，还体现在附带伤害问题上。现代战争伦理对于判定附带伤害的正义性存在两重标准：既认为因攻击合法的、关键的敌方目标而造成的附带伤害可以接受，又强调附带伤害必须受到严格的限制。而对于如何规制附带伤害，尚存有一定的争议。一种观点是从动机上分析，即必须证明战争对无辜平民的伤害只是偶然因素所致，而非出于故意。另一观点是从后果来分析，认为对无辜者的伤害或者是容易补偿的，或者被行动的后果所充分压倒。[①] 近年来，情报部门往往采取植入病毒的方式来获取反恐情报信息，其独特的“摆渡”攻击方式，进一步冲击了关于附带伤害的伦理规范。有学者认为，如果被攻击国家的关键基础设施的信息系统使用固定的互联网地址，就很容易被攻击者发现并准确定位。然而，现实并非如此。因此，有效的网络攻击必须要寻找目标并将攻击的范围扩大，这就会将感染病毒的危险转嫁给非战斗人员。[②] 面对与外部互联网处于物理隔离状态的工业控制网络系统，网

① 左高山. 战争镜像与伦理话语 [M]. 长沙：湖南大学出版社，2008：39.

② LIN P, ALLHOFF F, ROWE N C. War 2.0 cyberweapons and ethics [J]. Communications of the Acm, 2012 (3): 25.

络情报战必须以民用计算机系统为中介遂行攻击任务。从动机论的视角看，这种“摆渡”攻击造成的附带伤害，无疑是故意为之而非偶然因素，因而是有悖道义的。而从效果论的视角看，“摆渡”攻击造成的附带伤害是否微不足道且容易补偿，只能视具体的案例而定，难以统一评估，至于行动的后果能否充分压倒其所造成的附带伤害，攻击者与受害者显然会有不同的评判。

网络情报战引发的区分困境，还体现在情报战发起者在战争中的合法身份问题上。依据区分原则，战斗人员对非战斗人员的蓄意攻击是被禁止的，而不具备合法身份的人员参战同样也是不被允许的。网络情报战的出现与大规模实战应用，无疑造成了“战斗人员”概念的进一步泛化，任何具有操控网络攻防相关技能的个体与组织，都可能在网络反恐情报战中扮演“战斗员”“情报员”的角色。作为对战斗人员身份识别问题的一种回应，近年来，美国、英国、日本、俄罗斯、以色列等国开始将网络战力量列入军事编制序列。2009 年，美军在战略司令部下设立网络战司令部，将执行对外网络反恐情报战任务的网络“作战使命力量”纳入其中。当然，网络战司令部的成立并不意味着身份识别问题的终结。2011 年，美国国防高级研究计划局开展了“网络快速追踪”计划，让具备网络技能的小型公司和个人在短期内投入网络安全任务之中。[①] 任职于美国国防大学国家战略研究所的戴维等人认为，这一举措无疑拉近了美国军方与网络黑客之间的距离。[②] 网络黑客组织与美国中情局、国防部之

① DAWN L. DARPA's new "Fast Track" okays hacker projects in just seven days [EB/OL]. (2011 - 11 - 14) [2022 - 04 - 07]. http://www.wired.com/dangerroom/2011/11/darpa - fast - track/.

② DAVID K, PUDAS T, YOUNG B. Preparing the pipeline: The U.S. cyber workforce for the future [J]. Defense Horizons, 2012 (8): 5.

间的合作模式，可能会对区分原则构成潜在的挑战。无军事人员身份，不隶属于国防机构，甚至无网络执法权力的黑客群体，一旦执行具有国家意图的反恐情报战任务，必然与区分原则的伦理要求背道而驰。

在满足相称性原则方面，网络情报战的实施同样引发了质疑。依据相称性原则的理念，使用某种作战手段的道德成本应当同预期的道德收益相匹配。这就意味着攻击者必须控制武力使用的程度与范围，避免因超出军事目的所需而造成的过度杀伤。作为一种新兴的情报战手段，网络武器在情报搜集和网络作战中展现出极大的不可控性。由于网络病毒的传播依靠互联网和移动存储介质，它的感染范围与破坏程度并不完全取决于攻击者的意图与作战规划，而是在很大程度上受网络用户个人意志支配。正因为如此，网络情报战攻击的实际范围往往会超出既定的目标。例如，破坏伊朗核电站工业控制网络系统的“震网”蠕虫病毒，不仅感染了伊朗境内60%的计算机，还扩散到美国、英国、德国、澳大利亚、印度、印度尼西亚等国。虽然西门子公司声称蠕虫的扩散并没有造成实际的损害，[①]但考虑到所有 Windows 系统的计算机都需要进行杀毒软件的升级服务，以及全球使用西门子 WinCC 系统的工业设施用于相应安全防护的投入，这无疑是波及全球的一笔巨大损失。一项针对局部特定目标的网络信息窃取行为引发全球范围的网络危机，网络情报战与相称性原则的冲突可窥一斑。

网络情报战对相称性原则的挑战，同样体现在受害者如何对网络攻击施行报复。事实上，进攻方采取与军事目的匹配的攻击强度，

① MCMILLAN R. Stuxnet worm hit industrial systems [EB/OL]. (2010-09-14) [2022-04-07]. http://www.computerworld.com/s/article/9185419/Siemens_Stuxnet_worm_hit_industrial_systems?taxonomyId=142&pageNumber=1.

与被攻击方采取与所受损失匹配的反击强度，是相称性原则的一体两面。然而，由于当前网络情报战攻击效果的不可控性，被攻击方很难准确评估自身所受的实际损失，单一的病毒感染可能引发政治、经济、军事、文化多维度的连锁反应，进而导致被攻击方采取违背相称性原则的反击手段。

近年来，围绕“相称性原则”与“惩罚威慑”的研究，成为网络威慑理论研究的一大热点。部分学者认为，由于网络反恐情报战是双向的，国家与国家之间的信息产业发展水平以及网络依存度存在较大差异，遵循相称性原则就不能有效实现惩罚威慑，特别是网络基础设施不发达国家对美国发起的网络情报战，单纯在网络维度的反击，并不一定能发挥威慑效果。美国海军战争学院国际法系主任施密特等人认为，对对手的网络战进行惩罚报复，不应当限制在网络维度，而是可以考虑跨域威慑手段。乔治·华盛顿大学格拉泽教授则更进一步指出，应对恐怖主义国家发起的高度破坏性的网络攻击，美国可以以侵略对方领土作为威慑，甚至进行政权更迭。[①] 2011 年 5 月，白宫颁布了《网络空间国际战略：互联网世界的繁荣、安全和开放》，明确指出：“通过网络空间实施的某些敌对行为，将迫使美国根据对军事条约缔约伙伴的义务采取行动，美国保留根据国际法适当使用所有必要的外交、信息、军事和经济手段的权力。”这一网络空间战略的发布，既体现了美国捍卫网络空间安全的坚定决心，也体现了其推行以网络情报战为代表的网络攻防方式并没有绝对的伦理和法理限制，只要符合美国的国家利益，相关手段方法的使用就可以突破相称性原则的束缚，达到“无所不用其极”

① GLASER C L. Deterrence of cyber attacks and U. S. national security [EB/OL]. (2011 -06 -01) [2022 -04 -07] https://cspri.seas.gwu.edu/sites/g/files/zaxdzs5851/files/downloads/2011 -5_ cyber_ deterrence_ and_ security_ glaser_ 0. pdf.

的程度。由此可见，网络情报战看似只是应用于情报信息搜集，实质上已成为某些军事强国在网络空间谋求霸权的重要手段。

四、网络情报战与战后正义

在传统的正义战争理论中，虽然也存在战胜国与战败国之间，以及国家、集团、个人之间的责任分配困境，但战争的相关责任方是明确的。这是由于在陆地、海洋、天空、太空等实体空间，现有的技术手段基本可以锁定攻击的来源。然而，对于当前在网络空间的情报战而言，受害国家确定网络入侵的来源存在技术上的归因困境。美国国家科学研究委员会计算机科学与电信学部首席科学家认为，在以下五种情况下很难判定网络攻击的来源：一是网络攻击者使用前所未有的技术手段，如利用微软未曾公布的漏洞研制的计算机病毒；二是网络攻击者在入侵过程中没有留下恶意操作的证据；三是植入的病毒软件在完成搜集数据任务后自行毁灭，不留下任何访问痕迹；四是网络入侵行为没有明确的动机，或者行为不是发生在双边或多边冲突与对抗的氛围下；五是必须对网络攻击做出快速应对，阻碍了对入侵者进行深层次的调查。[①] 除了上述情况，考虑到网络空间存在全球公域与国家主权之争，跨国追查网络攻击的责任者可能还要面临网络司法管辖权的问题。技术、司法等因素造成的归因困难，使得网络情报窃取的责任者难以明确，网络入侵的责任分配难以确定，从而在技术层面上导致实施网络情报战难以被精准地追责。

① LIN H. Escalation dynamics and conflict termination in cyberspace [J]. Strategic Studies Quarterly, 2012 (3): 49 - 50.

虽然网络情报战的责任分配存在归因溯源的困难，但是，在某些学者看来，这并非完全无法解决的。他们针对网络入侵提出一系列的归因依据：以网络入侵的时间与国际关系背景结合来归因，以网络入侵手段的技术特征与国家信息技术力量结合来归因，以网络入侵的目标与相关领域的国际竞争结合来归因，以网络入侵的最大得益者来归因，以对待国际调查的态度来归因，拒绝接受调查的国家可以列为嫌疑对象，等等。[①] 上述归因依据虽具有一定的合理性，然而，考虑到目前网络空间仍处于霍布斯式的“自然状态”，主权国家并非实施网络入侵的唯一行动者，各式各样的黑客与恐怖组织同样可以匿名从事网络入侵活动，后两者的行为无疑具有更大的非理性成分，因而针对主权国家的归因依据对于后两类网络行为者并不适用。因此，很难将主权国家实施的网络情报战与黑客组织的网络入侵区分开来。例如，针对伊朗核设施控制系统的“震网”病毒，就无法确切归因于美国和以色列政府。在现行的 IPv4 网络标准下，归因技术一时难以突破，完备的归因标准体系也难以建立。对于网络入侵行为的错误归因，不仅无法实现惩治战争罪行的目的，而且可能导致国际冲突的升级与恶化，这无疑是与战争责任的伦理原则相悖的。

网络情报战对战后正义的冲击也体现在承担重建义务方面。依据人道主义原则，战胜国有维持战败国政治、经济和社会秩序的义务，战胜国和其他有能力的国家都应该对战后重建提供援助，以实现持久的和平。较之常规作战方式带来的“硬摧毁”或“地毯式的饱和轰炸”，网络情报战所造成的伤害在一定程度上具有可恢复性。一方面，当网络入侵造成数据丢失或程序崩溃时，运用原先的备份

① 何奇松. 美国网络威慑理论之争［J］. 国际政治研究，2013（2）：58－59.

就可以很快实现系统还原，这是热兵器造成的硬杀伤所无可比拟的。另一方面，诸如“火焰”病毒等恶意软件，一旦完成搜集情报数据任务就可自行毁灭，整个攻击过程甚至在不被察觉的情况下即已完成。基于上述两个方面的原因，一些伦理学家认为，关于网络攻击的损伤恢复并非网络情报战伦理研究的主要关注点。[①] 然而，也有学者对此提出异议，由于大数据时代网络系统的巨复杂性，可能会无法探知网络攻击的损伤，抑或网络攻击的损伤恢复需要耗费无法接受的漫长时间。[②] 有鉴于此，可尝试将“可恢复性”列入网络反恐情报战的伦理标准，作为规制网络攻击使用的方法和手段。实现网络空间的损伤恢复，应当成为彰显网络反恐情报战战后正义的可行性路径。

正义战争理论是在人类长期战争实践中逐渐形成与发展的，新的作战手段的应用破坏了旧的伦理原则，迫使战争伦理的理论体系做出相应的修正与完善，从而推动其内涵的不断丰富与外延的日益拓展，这个循环往复的过程被称为“重新商议”的伦理构建机制。围绕网络情报战的正义性，相关的伦理争议逐渐展开，“重新商议”机制又一次开启。从“商议”的意义上说，正义战争理论对于网络情报战的规制能否真正体现正义性和人道主义原则，在很大程度上取决于协商网络伦理规范的范围能否由少数强权国家扩展到整个国际社会（包括主权国家与非国家行为体）。正义战争理论究竟会以何种形式对网络情报战进行规制？现在看来尚未到论定之时。然而，基于网络空间攻易防难的脆弱性，一旦主权国家和非国家行为体都

① ROWE N. The ethics of cyberweapons in warfare [J]. International Journal of Military Ethics, 2010 (9): 384 - 410.

② LIN P, ALLHOFF F, ROWE N C. War 2.0 cyberweapons and ethics [J]. Communications of the Acm, 2012 (3): 26.

掌握了摧毁对方网络的技术手段，网络空间就可能出现类似核均势的“相互确保瘫痪”状态，这时实现对网络情报战的伦理规制和法律约束就大有可期。

第四章
分析决策的正义性：以大数据分析为例

在智能化战争的作战流程中，不仅需要在实体空间和虚拟网络空间搜集大量涉恐情报信息，而且需要对海量数据信息进行分析研判，为具体实施作战提供决策支持。大数据技术主导的分析决策，不仅是国家危机预警必不可少的环节，而且标志着国家安全战略决策朝着智能化完全取代人力方向迈出了一大步，进而对传统的正义战争伦理体系构成了一系列新的挑战。

本章阐述大数据分析的军事价值及应用，以及大数据分析面临的伦理困境，并有针对性地探讨“大数据分析如何合乎道义?”这个问题，为我国制定维护大数据安全的政策法规提供借鉴。

一、大数据分析的军事价值及应用

（一）大数据分析的军事价值

2019年8月9日，英国查塔姆研究所网站在题为Artificial Intelligence Prediction and Counterterrorism的报告中指出，大多数国家都专注于防止恐怖袭击，而不是被动对其做出反应。因此，预测已经成为有效反恐的核心。人工智能能够分析体量更大的数据，并且可以感知那些数据中的模式。[①] 由于网络空间成为恐怖分子与反恐国家之间进行安全博弈的重要战场，植根于网络空间的大数据情报信息，也日益成为维系国家公共安全的重要资源。在作战分析决策环节，以美国为代表的北约国家提出大数据研判理念，通过采取大数据技术对相关情报信息进行分析、加工处理，研判对方攻击行动规律，并为随之展开的反制行动提供决策参考和行动依据。

大数据在智能化战争中的价值日益凸显。有学者指出："在移动设备、网络、社交媒体、传感器、日志文件和事务性应用程序等来源的实时数据激增的情况下，大数据分析已经成为一种可以为当今战争提供丰富多样的关键任务数据的有效方法。企业用来挖掘数据库的大数据技术正被军方采用，从无人机、自动网络安全系统、恐怖分子数据库和许多其他来源的不同类型的数据中搜集更多信息。这项技术不仅有助于战场上的作战人员，还将用于改善从软件开发

① 安全内参. 查塔姆研究所报告：人工智能预测与反恐［EB/OL］.（2019-08-13）［2022-04-07］https://www.secrss.com/articles/12924.

到车辆维修等各个领域。”[①] 据统计，推特用户平均每天上传6.56亿条推文，脸谱用户平均每天发布43亿条消息，这是来自两个社交网络的数据，再加上谷歌每天的搜索量（52亿次）、优兔视频的观看量（每分钟400万次），相当于军方可获得的海量数据。[②] 一架MQ-9“收割者”无人机用于执行间谍任务，收集“相当于20台笔记本电脑”容量的数据。“收割者”和“捕食者”有12个摄像头，美国阿格斯地面监视系统每秒收集的信息超过40千兆字节。[③]

针对大数据领域日益凸显的军事价值，军事大数据分析悄然兴起。所谓军事大数据分析，就是依托大数据技术和分析工具，从海量情报数据信息中挖掘出隐藏其中的特殊关联性，使用关联分析、聚类分析、分类分析、异常分析、特异群组分析和演变分析等大数据分析方法，分析敌对国家、敌对组织的具体行为特征，勾勒敌对力量从事攻击行动的活动轨迹，提取敌对分子活动的时空规律，创建敌对力量网络活动特征的认知数据库。通过分析不同危机事件间的互动关系，构建敌对组织的行动网络。当前，利用特定算法对敌对危机行动的发生概率进行评估，研究和预判敌对袭击行动，已成为西方军事强国国家安全部门的通行做法。

（二）大数据分析在国家安全领域的应用

当前，大数据分析在国家安全领域的应用非常广泛。21世纪的

① KHAN A. Big gains foreseen for big data in military applications [EB/OL]. (2017-05-16) [2022-04-07]. https://www.defensemirror.com/feature/12/Big_Gains_Foreseen_For_Big_Data_In_Military_Applications.

② ABADICIO M. Big data in the military - preparing for AI [EB/OL]. (2019-05-08) [2022-04-07]. https://emerj.com/ai-sector-overviews/big-data-military/.

③ Analytics India Magzine. Using big data in military operations: this is how future wars will be fought [EB/OL]. (2015-10-03) [2022-04-07]. https://analyticsindiamag.com/using-big-data-in-military-operations-this-is-how-future-wars-will-be-fought/.

反恐作战，在军民混杂、高速流动的反恐战场，依靠服饰、武器等传统的标识性特征，难以确认恐怖分子的身份。有鉴于此，欧美军事强国在反恐作战过程中，更多地借助技术手段，特别是运用大数据技术和分析工具，深入把握恐怖分子和恐怖组织活动的路径模式与特点规律，为反恐作战提供强有力的信息支撑。概而言之，当前，大数据分析在国家安全领域的应用主要体现在三个方面。

其一，依靠大数据技术甄别、锁定敌对目标。大数据追踪技术是奥巴马政府2012年宣布的技术计划的核心。美国国防高级研究计划局制定了一项追踪大规模杀伤性武器威胁数据的国防部计划，将数据挖掘和专利分析结合起来，跟踪可能改变军事装备未来的颠覆性技术。[①] “9·11”事件之后，美国国家安全部门广泛利用大数据技术捍卫国家安全利益。美国国土安全部门实施国家情报网格计划，针对互联网络空间中异常活跃的涉恐网站、论坛和社交媒体，使用格鲁曼公司设计的大数据分析软件，进行涉恐敏感信息的大数据挖掘。在此基础上，有针对性地进行大数据聚类分析和关联分析，锁定活跃在社交网络的恐怖分子群体和组织。例如，2013年波士顿马拉松爆炸案后，美国国家安全部门就通过大数据分析网络搜索记录，将搜索过“高压锅”“炸弹”的网民列为恐怖袭击嫌疑人。同时，依据恐怖分子在网络空间的活动轨迹记录，勾勒其在现实世界生活的时空信息和活动轨迹，使用大数据演变分析，研判推演未来恐怖分子活动的潜在时空环境，为锁定目标提供信息支撑和辅助决策。

其二，运用大数据技术提高针对敌对国家、敌对势力渗透活动的网络态势感知。2014年，美国国防高级研究计划局宣布了一个名

① KHAN A. Big gains foreseen for big data in military applications [EB/OL]. (2017-05-16) [2022-04-07]. https://www.defensemirror.com/feature/12/Big_Gains_Foreseen_For_Big_Data_In_Military_Applications.

为“挖掘和理解软件飞地”的项目，以提高军方态势感知的能力。在锁定敌特分子、恐怖暴力事件嫌疑人、敌对组织特殊关联人等关键敏感目标个体的基础上，国家安全部门运用大数据技术把握敌对组织的整体行动态势，通过构建“敌对分子身份数据处理系统”的数据库收集储存相关数据信息，通过跟踪监控敌对分子开设的网络账号的信息发布、搜索热词和活动频率勾绘敌对组织网络活动的数据全息图。在军队，分析员必须能够过滤海量数据流，以识别和分类对军事战略和行动有任何用途或影响的信息。这可能与某些国家、特定个人、高危人群和武器等有关。[①] Voyager Lab 指出，虽然许多“独狼”攻击者是独立且不可预测的参与者，但当他们发展成为威胁时，通常仍会留下显著的、可识别的数字足迹。许多“独狼”恐怖分子通过在线平台分享想法、寻找支持的受众、表达仇恨并获取对其事业和使命至关重要的信息。这是人工智能探索和考虑的额外途径，因为它梳理了每天生成的公开可用的海量数据，这些数据需要专家和经验丰富的分析师团队花费无数小时来处理和分析。依托大数据分析工具，综合采用多源融合态势评估分析、威胁态势评估分析和态势预警等技术手段，掌控敌对组织向网站、论坛、社交媒体及移动客户端渗透的实时状态，以便在关键时间对关键节点（涉恐网站和涉恐社交账号等）做出及时响应，阻隔敌对分子的网络渗透渠道，切断敌对分子实施危机行动的指挥链路，配合实体空间军事打击，实现在线上与线下压制、瘫痪敌对组织的目的。

其三，以大数据模型预测恐怖袭击等危机事件的发生概率。国家安全部门不仅利用大数据应对当前出现的危机事件，而且致力于

① ABADICIO M. Big data in the military – preparing for AI [EB/OL]. (2019 – 05 – 08) [2022 – 04 – 07]. https://emerj.com/ai – sector – overviews/big – data – military/.

预判和评估敌对分子未来的行动趋势。有学者表示，美军利用大数据分析工具和技术，结合机器学习、文本挖掘及数据建模，搭建了敌对势力威胁预测、评估和早期预警的第一道防线。[①] 美军情报部门通过对危机事件政治诱因、民族诱因、宗教诱因，以及袭击时间与袭击位置分布等方面历史信息的大数据分析，揭示其中可演化的规律性信息。在此基础上，采取多步时间序列预测中的递推计算思想，对未来可能发生的危机事件进行预先评估，充分考虑每一步预测的不确定性，将其作为下一步预测迭代的输入要素进行模拟仿真计算，达到鉴定可疑行为模式、识别潜在危机事件威胁、预防和减少危机事件和犯罪活动发生的目的。

二、大数据分析的伦理挑战

较之传统战争模式，依托大数据技术的危机处理方式具有风险小、效费比高等优势，符合人工智能时代战争的轻量化诉求和智能化转向。然而，在作战决策过程中高度依赖大数据分析，引发了一系列的伦理道德问题，导致学界的激烈争论，引起国际社会的广泛关注。

（一）大数据分析对开战正义的挑战

大数据分析对开战正义的挑战，聚焦在正当理由层面，体现为以下三个具体的问题。

第一个问题，大数据分析是否具有正当理由？严格意义而言，只有对他国的侵略行为进行防卫，才是体现开战正义的唯一正当理

① SHAH K, DEZYRE. Big data and data science for security and fraud detection [EB/OL]. (2015-12-01) [2022-04-07]. https://www.kdnuggets.com/2015/12/big-data-science-security-fraud-detection.html.

由。然而，从17世纪开始，以国际法学家格劳秀斯为代表的部分学者，提出“杀死那些正准备杀人的人是合法的”，将对正当理由原则的解释由“自卫权”进一步拓展为“预先防卫权”。正当理由解释的拓展，对从开战正义层面审视大数据分析的正义性提出了挑战。对于已经发生的恐怖袭击，依托大数据技术锁定和打击恐怖分子和恐怖组织，其正义性是毋庸置疑的。但是，依托大数据分析技术预测可能发生的危机事件，实施“先发制人”式主动攻击是否具有正当理由？这就成为一个具有争议的伦理问题。一方面，对于那些确信无疑的、刻不容缓的、没有考虑余地的危机事件，获得大数据情报支持的反制行动属于本能反应式的自卫；另一方面，对经过大数据分析推演显示为遥远的、不确定的战争或者暴力袭击危险，如果一味滥用预先防卫权，将潜在威胁全部纳入发动预防性战争的凭据，就容易使少数霸权国家轻易逾越战争门槛，在战争与和平的抉择中不断挑战更低的道德界限。

第二个问题，基于大数据分析的判断决策，在多大程度上可以成为发起自卫作战的凭据？如果一国凭借大数据分析预判提供的证据行使预先防卫权，进而对潜在敌对目标发动先发制人的攻击，那么需要满足的一个必要条件是，基于大数据挖掘和分析得出的情报信息在道义上必须足以支撑行使预先防卫权的需要。依据开战正义的正当理由原则，发动正义战争应当提供明晰而准确的是非判断。具体而言，作为开战的正当理由，必须能证明潜在的战争或恐怖袭击事件确实会发生且国家安全确实会受到侵害。然而，大数据挖掘情报信息的过程，即对海量数据实施情报检索、统计分析、评估预测的过程，其对潜在袭击事件发生概率的预判存在诸多不确定性：一是情报数据的选取、分类环节，选取涉恐数据的标准、数据分类模型的参数设置和模型构建等因素，可能在源头上影响大数据分析

的准确性；二是数据的估值环节，如何给定输入数据以及如何预设算法进行估值，都会影响输出端得到未知连续变量的值；三是潜在战争或袭击事件的预测环节，如何建立模型和假设、如何测试和验证挖掘结果以及如何处理错误或不一致的运算结果，都会影响预判的准确性，进而左右先发制人作战理由的正当性。

第三个问题，大数据情报分析对他国数据信息的侵犯，是否违背开战理由的正当性与合法性？近年来，随着美国信息安全战略逐渐实现“由守转攻”的演变，在网络空间利用大数据分析挖掘手段，发现潜在安全隐患，有针对性地实施预防性反恐战略，日益成为美国政府捍卫国家安全的重要做法。由美国中情局前雇员斯诺登披露的“棱镜门”事件，就是美国在他国网络空间挖掘数据、搜集情报的典型案例，由此带来的战争伦理问题，即大数据情报分析是否构成对他国数据信息的侵犯。近年来，美国为了维护国家安全，往往通过网络黑客窃取涉及他国国家安全的情报信息，进行大数据挖掘和分析研判，这无疑导致依靠大数据分析发动战争在合法性方面存在巨大的争议。

（二）大数据分析对交战正义的挑战

交战正义特别关注战争手段的运用应当符合必要的道德准则。大数据分析手段在军事行动中的广泛应用，虽然在一定程度上提升了锁定敌对国家和预判潜在危机事件的效率，但也导致在作战行动中的“技治主义”倾向。片面强调大数据手段和数据量化标准，难免引发作战行动过程中的一系列伦理问题。

首先是对区分原则的挑战。作为交战正义的核心内容，区分原则主要关注交战过程中攻击对象的合法性问题。具体而言，就是指攻击行为要坚持“非战斗人员豁免权”，在军人与平民、军事目标与

民用目标之间作出严格的区分。大数据分析研判的军事应用，从以下三个方面对区分原则构成冲击。

一是偏重可量化的区分依据。大数据分析研判模式下，国家安全机构对于合法目标的锁定与区分，主要基于日志文档、电子表格、数据库等结构化数据，而图片、音频、视频等非结构化数据，以及其他无法简单量化的信息，则无法通过大数据平台转化为判别敌对分子的依据。对此，美国学者布特指出，美军过分看重数据的区分方式，导致判定阿富汗塔利班恐怖分子困难重重，而当地农民则可以一眼看出谁是塔利班。这种差异性是由于大数据技术难以实现阅历、生活经验和主观感受的数据化，因此，最先进的大数据分析技术也难以有效研判难以量化的目标对象。①

二是营造虚拟化的区分环境。得益于大数据平台的支撑，作战人员可以在远离战场的虚拟环境下遥控实施作战行动。然而，他们对于战场的感知是基于视频设备和海量数据。依托大数据构建的虚拟化区分环境，使得数据分析人员感受不到真实战场的氛围，造成“重数据不重人、见数据不见人”的意识倾向，进而对剥夺他人生命产生麻木心态。

三是唯数据主义的火力打击模式。所谓唯数据主义的火力打击，就是在犯罪嫌疑人真实身份尚未确认的情况下，利用大数据分析挖掘技术，对符合某些预设嫌疑分子数据信息特征的目标实施先发制人行动。近年来，美军在阿富汗、巴基斯坦、也门等地人力情报资源匮乏的情况下，单纯依靠大数据技术来预判。一旦通过分析认定某人具有与恐怖行动相匹配的数据信息特征，就在未经国会授权和

① 布特. 战争改变历史：1500 年以来的军事技术、战争及历史进程［M］. 石祥，译. 上海：上海科学技术文献出版社，2011：392.

未经确认身份的情况下实施无人机“定点清除”行动。这种唯数据主义的打击模式，招致国际社会和媒体舆论的广泛批评。

其次是对相称性原则的挑战。与区分原则一样，相称性原则也是交战正义的重要体现。依据相称性原则，交战双方必须控制武力手段运用的范围与程度，避免出现超出军事目的所需的过度杀伤。由人工智能平台主导的军事打击行动，依靠智能化指挥系统和“全球信息栅格”提供的精准坐标信息，用挂载微型化精确制导炸弹的智能化作战平台实施远程精确打击，原本可以作为遵循相称性原则的典范。但是，过度依赖大数据技术难免造成相称性原则在军事作战行动中发生异化，即片面追求数据层面的相称，忽视道德价值上的相称。相称性原则作为正义战争理论的原则，其本质是追求道德成本和道德收益的相称，大数据分析研判将这一原则片面数据化，以数据层面的相称性替代道德价值层面的相称性，将击毙敌对分子的数量与打击效率等量齐观。在恐怖组织头目数量有限且难以锁定的情况下，美军频繁使用价值百万美元的“地狱火”导弹清除底层恐怖分子，无疑违背了正义战争理论中相称性原则的道义初衷。更值得关注的是，大数据分析研判在智能化战争中的应用，体现出西方军事强国在作战中急功近利的功利主义思想，即片面追求短期的数据相称，忽视长期的和平安全。通过大数据技术高效率的锁定目标能力，实施高频次的打击行动，虽然短期内能够实现区域安全治理，但长期来看，可能导致区域内民众的恐慌情绪不断累积，甚至可能造成极端民族主义、恐怖主义活动愈演愈烈。

三、大数据分析如何合乎道义？

大数据技术在智能化战争中暴露的道德困境，凸显其在伦理规

约层面的失范状况。围绕大数据技术军事应用构建伦理道德规范，必须将其贯穿开战正义、交战正义和战后正义的全部过程。

首先，依靠大数据技术分析研判并行使预先防卫权的必要条件，即确保足够充分的正当理由，其核心在于证明一国遭遇战争或恐怖主义袭击威胁的充分程度，具体体现为以下三个方面。第一，公开表明实施袭击的意图，运用大数据技术研判敌对国家或恐怖组织的行动意图，不应根据潜在的或猜测的情况，而应根据公开表述或已摆出袭击态势的情况。第二，敌对国家或恐怖组织的积极准备已经使袭击意图变成现实危险，作为预先防卫方，国家安全部门必须提供敌对国家或恐怖组织的准备工作造成实质伤害的有力证据。第三，坐视不顾将加剧国家安全威胁的风险。依据开战正义的最后手段原则，只有当面临潜在袭击的威胁逐渐累积，由量变转化为质变，且其他非暴力手段无法应对的情况下，军事行动作为先发制人的手段才具有正当性。

其次，大数据技术军事应用需要有效遵循区分原则和相称性原则。在遵循区分原则方面，应当坚持数据特征区分与非数据特征区分并重。一方面，保留虚拟环境下必要数据区分手段，要继续发挥大数据技术在识别安全风险和锁定敌对分子嫌疑人方面的效率优势；另一方面，鉴于大数据主导的区分方式存在诸多弊端，还必须重视真实战场环境下的直接感官体验，将其作为识别战争威胁不可或缺的重要凭据，打破区分权集中在大数据分析专家手中的现状，充分尊重一线作战官兵和当地民众的经验判断，充分尊重真实战场感官在目标搜索、识别和判断中的参考价值，实现大数据技术与感官经验的融合。在遵循相称性原则方面，要矫正大数据分析研判模式下唯数据化的相称理念。一方面，注重非数据层面的相称性，不能简单以锁定和击毙敌对分子嫌疑人的数量来衡量军事行动的正义性，

规避对潜在嫌疑人一视同仁的“斩首”模式，而应依据敌对分子级别和威胁紧迫程度，采取差别化的应对方式；另一方面，大数据研判应更关注长时段的相称性，妥善处理短期效率和长期持久和平的关系，不以作战数据作为衡量军事行动成败的唯一标准。例如，在反恐行动中，应当着力从根源上铲除恐怖主义，从谋取反恐区域内原生居民的福祉出发，保留和解与恢复战前活动的可能性，避免造成比战斗更长久的痛苦的风险。

最后，鉴于经典正义战争理论对战争责任的界定难以适应大数据技术军事应用的现实挑战，应当进一步细化对战后正义的伦理规约。第一，主体上凸显人的责任。坚持大数据技术的责任主体是人而非智能化设备的理念，明确大数据分析技术只是手段，只能简化反恐过程和提升作战效率，坚定军事指挥官作为作战行动最终决策者和主体责任人的地位。第二，细化每一个环节的责任分配。按照大数据技术应用的具体环节细化责任分配：在数据获取环节，厘清原始数据搜集和大数据挖掘之间的责任分配；在数据分析环节，明确客观数据呈现和作战人员主观判断之间的责任分配；在研判决策环节，明确军事指挥官误判和技术保障人员失误之间的责任分配。第三，出台规避责任转嫁的政策法规。一方面，为避免国家以所谓大数据技术失误逃避战争责任，应当出台相应法律法规，明确军方和国家安全部门具有公布大数据技术运用细节的责任，让其接受民众和国际机构的监督；另一方面，针对潜在战略对手利用大数据技术实施嫁祸行为，应在联合国框架下建立大数据技术应用调查机构，构建联合核查机制，明确对第三方介入的追责与反制措施。

第五章
无人机“定点清除”行动

近年来，数字信息网络、大数据、人工智能技术的发展日新月异，在社会生活领域和战争领域中都有广泛的应用。在社会生活领域的应用造成了如信息鸿沟、算法歧视、大数据杀熟等一系列伦理道德问题；在战争领域的应用则直接催生了无人机作战这一全新的作战模式，对既有的战争伦理理论和原则造成了冲击与挑战。如何深入系统地揭示无人机作战的伦理问题并提出相应的规制路径，成为本章研究的一个重要考量。

无人机“定点清除”行动是21世纪以来美军在全球范围内打击恐怖分子所采取的一种重要的作战方式。本章重点分析美国无人机“定点清除”行动的正义性问题。

一、美国无人机“定点清除”行动的发展

美国是执行无人机“定点清除”行动最为频繁的国家。美国自小布什政府开始，大力推动无人机执行反恐作战任务，并将反恐战略由大规模常规战争的全面反恐调整为主要依托无人机实施的“点穴”式反恐，不断扩大无人机反恐的范围。自阿富汗战争以来，美国构建了以内华达州克里奇空军基地为核心的无人机反恐作战指挥中心，远程指挥和控制遍布全球的无人机进行针对恐怖组织头目的“定点清除”攻击。美国中情局和美国空军在海外执行反恐任务的无人机都是由这一基地进行远程操控。[①] 在奥巴马执政时期，美国政府利用无人机进行反恐作战的频率日益增长，无人机已经成为集情报侦察、目标搜索、火力打击于一体的智能化反恐作战平台。

从作战范围来看，美国空军和中情局控制的“察打一体”无人机在全球以“反恐”为名频繁执行暗杀任务，其活动踪迹已经突破阿富汗、伊拉克等美国和北约驻军维和区域，日益覆盖巴基斯坦、叙利亚、也门和索马里等主权国家的空域范围。2015 年 5 月 7 日，美国对也门发动的无人机攻击，击毙“基地”组织阿拉伯半岛分支组织头目之一安西。该头目曾宣布对 1 月发生在巴黎的《查理周刊》恐怖袭击事件负责。针对身着便服、行踪飘忽不定的恐怖分子，无人机能够深入险境，第一时间将其锁定并击毙，因而具有常规反恐力量难以比拟的作战优势。

特朗普执政时期，日益倚重无人机这一低成本、高效率的反恐

① CBS NEWS. Behind the scenes of the Air Force's anti - terrorism drone program [EB/OL]. (2019 - 07 - 23) [2022 - 04 - 07]. https://www.cbsnews.com/news/air - force - anti - terrorism - drone - program - behind - the - scenes/.

利器，在美国发动的全球“反恐战争”中，无人机作战的应用范围和规模不断扩大，美国军方、中情局、国土安全部都参与其中，构成了美国全球无人反恐作战的庞大网络。特朗普任期内，在美国本土基地的远程指挥下，美国无人机在全球反恐行动中取得了一系列辉煌战果。2017 年 10 月 19 日，美军无人机空袭阿富汗与巴基斯坦边境地区，恐怖组织“巴基斯坦塔利班”分支机构“自由者大会”的头目霍拉萨尼被炸伤，随后伤重不治死亡。2018 年 12 月 2 日，以美国为首的联军向外界透露，联军在叙利亚的一次无人机“斩首”行动中，击毙了极端恐怖组织“伊斯兰国”高级头目乌马拉恩。2019 年 10 月 31 日，美国五角大楼公布视频和图像资料，宣布美军对恐怖组织“伊斯兰国”头目巴格达迪进行了成功的“斩首”行动，彻底切断了“伊斯兰国”的指挥中枢。2020 年 1 月 3 日，美国总统特朗普发表讲话，对外宣称美军运用“收割者”无人机成功击毙了伊朗“圣城旅”指挥官苏莱曼尼。这一事件再一次震惊中东乃至全世界，引发全球广泛争议。

据数据资料统计，截至 2019 年 3 月 1 日，美军用于反恐“定点清除”的 MQ－1“捕食者”和 MQ－9“收割者”无人机总飞行时数突破 400 万，它们能够全天 24 小时在全世界范围内连续执行侦察、搜索、救援和攻击及多种战斗和指挥任务。①

无人机在“定点清除”行动中的成功应用，进一步推动了智能化作战技术的创新发展。尽管当前无人机技术的发展距离真正意义上的智能化还有一定距离，“人在回路”地面操控模式仍然在主导无人机技术的发展，但是在“定点清除”行动使用的无人机系统中，

① 航空工业信息网. 美国空军 MQ－1 和 MQ－9 察打一体无人机总飞行时数突破 400 万 [EB/OL]. (2019－03－13) [2022－04－07]. https://www.81uav.cn/uav－news/201903/12/53483.html.

智能感知技术、机器学习技术、嵌入式计算技术、智能分层控制技术等智能化技术已经被大规模运用，用来突破操控人员生理因素的限制，大幅度提高无人作战系统在复杂的战场环境中的目标感知能力、自组织能力和自主控制能力。依据无人机作战的流程，当前智能化技术的应用主要体现在三个方面。

一是智能识别。目前，各国军方在无人机“定点清除”行动的识别区分中，已经尝试利用智能化技术来识别作战目标，为作战指挥呈现动态、可视的战场环境。智能化技术主要包括仿生视觉识别技术、基于认知的智能学习与推演技术、多传感器信息融合的战场态势评估技术等，主要涉及无人作战系统的识别区分过程中计算机的深度认知与学习、对复杂战场环境下的多元化目标进行实时感知和区分。通过功能强大的智能识别系统，无人机操控员不仅可以掌握战场气象、水文、地质等自然环境信息，而且可以感知作战目标的力量分布、火力配置、武器装备和运动的状态，为操控员进行对敌协同作战决策提供信息支持，以实现精确锁定作战目标，以及实时最优地更新作战目标活动轨迹，预测运动趋势。

二是智能控制。目前，美军将智能控制技术应用于反恐的无人机，无论是“全球鹰”无人侦察机，还是“捕食者”“收割者”无人攻击机，都是采用“平台无人、系统有人”的远程控制体系。虽然尚未实现无人机自主控制，但是已经在无人机搜索与跟踪控制、无人机自主飞行控制、无人机武器发射控制，以及无人机故障预测与自修复控制等方面取得突破性的进展。同时，通过可变自主权限的自主控制技术、快速动态任务重新规划技术等智能控制技术创新，已经初步实现信息挖掘、信息分析和为控制中心提供辅助决策的功能，在突发状况下辅助无人机操控员进行动态任务规划，提高无人作战系统作战响应的时敏性。

三是智能攻击。执行层级方面的智能化是无人作战系统应用于实战的基础。真正意义上的自主攻击，要求无人作战系统在不依赖外界指令和设备支持的情况下，在复杂作战环境中依靠机载智能系统独立完成攻击任务。目前，虽然无人机远程控制中心和操控员仍然掌握着执行攻击的最后按钮，但是美国和其他一些世界军事强国已经开始研制具备自主开火权的智能化机器人。由此可以预见，随着未来人工智能技术军事应用程度的不断深入，实现智能化自主攻击将成为军用无人机发展的一个重要方向。

值得注意的是，执行无人机“定点清除”行动的不是单一的智能化无人作战平台，而是一个集情报搜集、数据分析、作战决策和火力打击于一体的综合化作战系统。除了用无人机完成火力打击任务，在情报搜集方面，各主要军事强国还纷纷构建了覆盖全球的情报信息搜集网络。其中，以美国“棱镜”计划为代表的网络情报系统为美军当前的反恐作战提供了强有力的信息支撑。在数据分析和作战决策方面，全球的情报网络和信息化的战场环境导致战场数据呈几何级数递增，加之“反恐战争”节奏的不断加快，如何提升情报数据处理的时效性，确保对战场态势的实时感知与快速应对，成为“定点清除”行动战场指挥官高度关注的问题。针对这一问题，美国建立了以云计算、大数据分析技术为基础的标准化反恐信息分析、研判和决策系统。通过建立云计算中心来合并其下辖的几百个独立的信息中心，在全球范围内运用云计算技术实现情报信息的共享与共同处理。在基于云计算的数据处理环境中，各类传感器搜集到的敌对分子数据信息通过网络汇总到国家安全部门信息中心的“云端”存储系统，依靠功能强大的云计算中心和服务器组承担全部的计算任务，能够在瞬间完成对超大规模数据流量的处理。同时，利用大数据分析技术，军方能够通过数据计算实现实时快速的智能

化战场情报融合、战场态势实时感知和威胁等级实时评估等，提前预判潜在作战目标的行动趋势，并未雨绸缪提出预防性措施。

未来，无人机“定点清除”行动将向着真正自主化、无人化的方向发展。目前，无人机仍然由人类远程操控，执行“定点清除”任务的无人机操控员需要具有超出常人的作战能力，面临强大的生理和心理压力。例如，在虚拟的战场环境下，他们必须具有超人的专注度，及时通过指挥中心屏幕上的画面，感知战场态势变化。对无人机“定点清除”行动需求的进一步增长，促进集自主探测、识别、决断和作战功能于一体的自主化无人作战系统的研发。美军在《美国无人机发展路线图 2005—2030》中，进一步明确提出了无人作战飞机智能化自主作战技术，将无人机的智能化程度划分为由低到高的10个等级，计划由“平台无人、系统有人”的遥控操作模式，逐步发展为“平台无人、系统无人”、完全由计算机自主控制的“自律”无人机。由此可以预见，自主化将成为无人机“定点清除”行动未来发展的一个重要趋势，无人机“定点清除”行动未来可能在整个作战过程中彻底脱离人的控制。

作为智能化技术应用于军事领域的最新进展之一，无人机“定点清除”行动虽然具有广阔的军事前景和独特的性能优势，但制约其与军事部门深度融合的因素仍然存在。美国等西方国家利用“定点清除”行动造成的伦理危机和人道主义灾难，成为全球舆论和国际社会的热点话题。美军长期在阿富汗、巴基斯坦、伊拉克、索马里、也门等主权国家领空范围内执行无人机“定点清除”任务，且殃及大量无辜平民的生命。无人机“定点清除”行动的正义性与合法性问题，已经引发了伦理学界和法学界的广泛讨论。

二、无人机“定点清除”行动正义性问题的研究综述

伴随战争形态向智能化战争转变，无人机“定点清除”行动日益成为人类军事对抗的新兴作战方式，日新月异的攻防手段和作战方式创新，对机械化战争时代形成的战争伦理体系进一步构成强有力冲击，引发学界对无人机“定点清除”行动正义性问题的持续关注。本章在梳理国内外学界关于无人机“定点清除”行动正义性问题研究成果的基础上，总结这一领域的研究热点，揭示当前研究的薄弱环节并展望未来的发展趋势。

（一）无人机“定点清除”行动伦理研究的滥觞

2001 年，阿富汗战争爆发后，美军开始探索使用“察打一体”的无人机以“定点清除”方式攻击塔利班军事力量和“基地”组织恐怖分子，这成为无人机应用于“定点清除”行动的滥觞，由此也开启了围绕无人机“定点清除”行动正义性的讨论。在早期的研究成果中，克劳利和斯维塔克等学者早在 2002 年就开始关注使用无人驾驶飞行器引发的道德问题。[①] 由于当时无人机在战争中主要扮演辅助角色，因此，在阿富汗战争和伊拉克战争期间，探讨无人机“定点清除”行动正义性问题的研究成果并不丰富，相关研究具体呈现以下几个特点。

首先，学界尚未将无人机“定点清除”行动作为一种独立的战争样式进行伦理探讨。在 21 世纪初发生的阿富汗战争和伊拉克战争

① VINCE C, AMY S. UAV strike raises moral questions [J]. Air Force Times, 2002 (17): 16.

中，无人机并不是战争的主体力量，围绕无人机或无人驾驶飞行器的战争伦理研究，主要是在广义的机器人伦理的大框架下展开，具体的研究成果主要分为两类。一类是从自动化乃至初级智能化技术的宽泛视域出发，探讨自动化设备在经济社会生活不同领域广泛应用引发的伦理道德问题。例如，在桑蒂斯等人发表的一篇关于人机交互的伦理研究论文中，围绕社会不同领域中涌现的大量人机交互问题，探讨了智能化机器人介入人类生活后对传统伦理规范的冲击，其研究的重心为机器人在医疗、金融、商业和服务领域人机交互的伦理规范，而不是仅仅局限在军事领域。[①] 另一类研究虽然将研究视阈限定在军事领域之内，但往往将无人机置于整个军用自动化武器装备的大框架之内展开伦理探讨，而非聚焦无人机反恐的专题研究。例如，盖特林2005年发表的题为《致命性自主武器——伦理上和学说上的意义》的研究报告中，就将军用无人机与军用机器人及其他无人武器系统并列，作为信息化战争范畴下的一类作战手段，而未在独立的智能化战争语境下探讨相关的伦理问题。[②]

其次，在无人机应用于“定点清除”行动的初期阶段，学界已经开始关注无人作战手段对现行战争伦理规约的冲击。由于沃尔泽等人构建的经典正义战争理论体系，主要是立足于机械化时代的战争实践，难以适应对智能化战争进行伦理规约的现实需求。因此，自美国在阿富汗战争中使用无人机“定点清除”以来，对无人作战特别是无人机攻击手段进行伦理探讨已引发学界的关注。当时，学

① SANTIS A D, SICILIANO B, VILLANI L. Research on human - robot interaction: physical issues, with cognitive and ethical aspects [J]. Seminias in Laparosopic Surgery, 2002 (4): 230 - 243.

② GUETLEIN M A. Lethal autonomous weapons—ethical and doctrinal implications [R]. Naval War College, 2005.

界主要从无人机“定点清除”行动的军事实践出发，探讨无人机“定点清除”行动对交战正义的区分原则和相称性原则的冲击。这方面具有代表性的文章，当属科尔夫2003年在《军事伦理杂志》上发表的一篇探讨美军在也门进行无人机“定点清除”行动的伦理研究论文。作者敏锐地洞悉了无人机这种新兴的反恐手段对经典战争伦理体系的冲击，并结合美军在也门使用无人机“定点清除”行动的案例，分析了无人机反恐产生的伦理问题：“在遵守战争伦理和国际法方面，无人机操控员缺乏专业化的无人作战规范训练，仅仅通过从父母那里获得的规范、他们的文化和宗教信仰、学历等来决定攻击是否符合道德规范，而程序员通过对无人机控制系统进行适当的编程向其提供道德规范。”① 在科尔夫看来，程序化道德预置导致了无人机在反恐作战中遵循区分原则方面所存在的一系列问题。此外，麦考密克也在《无人机攻击》一文中，对美国无人机“定点清除”行动可能造成潜在的相称性伦理问题进行了研判。他认为无人机杀死敌人不会造成己方人员伤亡的特点，可能会诱发更为频繁的杀人行为，无人机飞行员和士兵学会了通过远程控制杀人，并且无人机将替代人类承担大部分的战争风险，将会导致人类失去对战争的恐惧意识。② 总体而言，当时的研究虽然触及了美国无人机“定点清除”行动的部分伦理问题，但观点相对比较零散，缺乏系统性、深入性的研究成果。

最后，学界对于无人机“定点清除”行动的伦理思考尚处于初级阶段，某些学者的观点存在一定的极端性和片面性，主要体现为

① KOLFF D W. Missile strike carried out with Yemeni cooperation—using UCAVs to kill alleged terrorists: a professional approach to the normative bases of military ethics [J]. Journal of Military Ethics, 2003, 2 (3): 240 - 244.

② MCCORMICK P. Attack of the drones [J]. U. S. Catholic, 2005 (8): 46 - 48.

两个方面。一类观点认为，既有的战争伦理体系是为了约束人类在战争实践中的行为而构建的，由于无人机“定点清除”行动没有人类直接参与，所以无人机战争可以彻底废除既有的战斗规则。[①] 此类观点存在的问题在于只看到无人作战“平台无人”的表象，却忽视了整个无人作战系统仍然由人类掌控的事实，因而做出无人机反恐可以摆脱伦理规约束缚的谬断。另一类观点认为，无人机“定点清除”行动的作战方式是精确制导弹药和远程瞄准技术持续发展的产物，无人机“定点清除”行动与传统的信息化战争条件下的作战方式没有本质区别，因此无须进行专题的伦理探讨。[②] 此类观点将无人机“定点清除”行动与信息化战争的作战方式混为一谈，与当时无人机的发展水平及其在反恐作战中的应用情况密切关联。在21世纪初的阿富汗战争和伊拉克战争期间，无人机作战在信息化战争中处于从属地位，无人机应用于“定点清除”行动尚处于探索阶段，其作战实践的数量非常有限，纯粹意义上的无人机作战尚未出现，其潜在的伦理风险尚未凸显，因此造成对无人机“定点清除”行动的伦理认知偏差。

（二）无人机“定点清除”行动正义性问题研究

伊拉克战争结束以后，伴随无人机作战技术的飞速发展，其在作战实践中的应用日益丰富，为推动无人机“定点清除”行动成为独立的战争样式创造了条件。在侦察预警、空天作战和远程打击行动中，各类智能化无人作战平台被大规模使用，其中，无人机在

---

① LEKEA J K. Missile strike carried out with Yemeni cooperation—the war against terrorism: a different kind of war? [J]. Journal of Military Ethics, 2003 (3): 230-239.

② DINAUER S, SWEENEY P. The joint force commander's ethical challenges for the 21st Century [R]. Naval War College, 2006.

“定点清除”行动中的应用尤为普遍。无人机“定点清除”行动日益由辅助作战手段演变为独立的作战手段。

使用无人机执行“定点清除”行动，不但对人类战争形态、战争样式的发展演进产生革命性影响，而且对现代战争伦理体系构成强烈冲击。对无人机“定点清除”行动的研究，已经突破信息化战争的研究视阈局限，无人机“定点清除”行动及其所代表的人工智能技术的军事应用，成为科学技术冲击现代战争伦理体系的典型案例。伦理学界开始独立、系统地探讨无人机“定点清除”行动的正义性问题，近年来国内外出版的一些高水平研究著作具体分析无人机“定点清除”行动引发的一系列伦理风险，相关研究涉及伦理、法理、制度、政策和战略战术等多个维度，主要围绕以下几个问题展开。

1. 无人机“定点清除”行动与开战正义

学界主要围绕正当理由、合法权威和正当目的三个伦理原则，具体探讨无人机“定点清除”行动对正义战争理论体系中开战正义原则的冲击。

首先，探讨无人机“定点清除”行动对正当理由原则的挑战。正当理由原则是维系开战正义最为核心的条件。在开战正义理论框架下，无人机“定点清除”行动伦理研究的首要问题是探讨无人机“定点清除”行动的正当理由。当前，学界主要围绕行使自卫权和预先防卫权这两种不同类型的理由，具体探讨无人机“定点清除”行动在开战理由方面的正当性问题。

一类是围绕行使自卫权的争议。依据正义战争理论，自卫是战争权利规约的唯一正当理由。因此，目前学界关于无人机“定点清除”行动是否符合正当理由原则的讨论，主要围绕行使自卫权展开。美国政府、军方和部分西方学者扩大对“自卫”的解释，认为美国

无人机作战是对发动“9·11”事件的“基地”组织恐怖分子行使自卫权，因此具有正当性。钱铖和石斌在系统研究美国政府对于无人机作战正当理由的一系列陈述后指出：美国政府认为，阿富汗塔利班政权庇护下的“基地”组织不但导演了“9·11”事件，也在此前此后策划实施了大量针对美国的恐怖袭击，对美国安全造成侵害并持续构成严重威胁。美国对“基地”组织和塔利班的打击属于自卫权的延伸，从而给予推翻阿富汗塔利班政权的军事行动以正当的理由。而“基地”组织在巴基斯坦和也门等国重新获得立足之地后，这一理由仍然有效。[①] 然而，许多学者对于这一行使自卫权的“正当理由”提出质疑，认为国家自卫行为应当发生在遭遇侵略之后，必须以国家主权受到根本上的侵犯作为前提条件，防卫权的行使必须是在领土、领海、领空等国家主权范围内。例如，艾哈迈德指出：美国在诸如巴基斯坦这样的非战斗区域和主权国家领空范围内实施无人机反恐计划，本质上是对所在国家主权的侵犯，由此可能引起一系列的法律和道德问题，最终将导致外交政策军事化和不必要的冲突。[②]围绕这一问题，黄云松和蔡瑞艳从国际武装冲突法的视角指出：在巴基斯坦、也门、索马里等国并非美国交战对象的前提下，美国在其领空范围内使用无人机反恐必须满足两个额外的条件。一是恐怖分子的活动导致上述国家已经失去在反恐区域内领土、领空的控制权，无法阻止恐怖组织从事危害美国国家安全的行为；二是美国在上述国家使用无人机“定点清除”行动得到了所在国政府的官方授权。在黄云松、蔡瑞艳看来，当前美国的无人机反恐未能满

① 钱铖，石斌.“雄蜂”的阴影：美国无人机作战对当代战争伦理的挑战［J］. 世界经济与政治，2013（8）：86－99.

② AHMAD M. The use of drones in Pakistan: an inquiry into the ethical and legal issues［J］. Political Quarterly, 2014（1）: 65－74.

足以上两个条件中的任意一项，因此不具有正当性与合法性。①

另一类是围绕行使预先防卫权的争议。具体而言，学界形成这一争议的背景是人工智能技术在“反恐战争”中的应用，它为预判潜在的恐怖主义威胁并实施预先防卫提供了技术手段。“9·11”事件以来，美国等无人作战技术先发国家构建了“棱镜”等智能化大数据分析平台，运用大数据挖掘、模拟仿真、兵棋推演等技术手段预判可能存在的国家安全威胁，以此作为发动“先发制人”军事行动的凭据。在此形势下，博伊尔指出，在奥巴马执政时期，美国无人机“定点清除”行动在法律和道德层面出现新的转变，集中体现为依靠人工智能技术预先研判、推定恐怖主义威胁，并借此行使预先防卫权，在可能存在“基地”组织及其盟友的世界上任何地方进行袭击，从而降低使用武力的传统障碍。② 施瓦兹将美国先发制人的无人机“定点清除”行动比喻成一种“治疗或预防政治癌症的医学手段”，可作为预防性治疗恐怖主义顽疾的“外科手术刀”使用。③ 博伊尔和施瓦兹等人所提出的这种基于预先防卫权的开战理由，对武装冲突的传统法律和道德标准提出了根本性的挑战，由此也引发了学界的激烈讨论。沃斯特对中情局使用无人飞行器“定点清除”恐怖分子和叛乱分子嫌疑人的案例进行道德评估，针锋相对地指出，美国以预先防卫为由实施先发制人的无人机“定点清除”行动，违背了善良、诚实、仁慈、勇气和审慎等基本道德标准。沃斯特特别批判了美国以“特征攻击”方式（即具有想定中恐怖分子的某些特

① 黄云松，蔡瑞艳. 无人攻击机面临的国际法挑战：以美国在巴基斯坦的无人机攻击为例 [J]. 南亚研究季刊，2012 (1)：21－25.

② BOYLE M J. The legal and ethical implications of drone warfare [J]. The International Journal of Human Rights, 2015 (2)：105－126.

③ SCHWARZ E. Prescription drones：on the techno－biopolitical regimes of contemporary “ethical killing” [J]. Security Dialogue, 2016 (1)：59－75.

征即可直接击杀）实施无人机反恐的行为，指出这种无须审讯即杀人的方式违背了伦理道德的基本原则。①

其次，探讨无人机“定点清除”行动对合法权威原则的挑战。作为开战正义的重要原则，合法权威原则规定战争是国家与国家之间的行为。② 在传统战争模式下，主权国家和联合国是发动正义战争的合法权威，无人机“定点清除”行动对合法权威原则的挑战主要体现在两个方面。一方面，无人机独特的隐秘攻击方式使得支撑开战合法权威的程序正义形同虚设。当前学界从美国无人机“定点清除”行动的不同战争案例出发，具体探讨其对合法权威原则的挑战。伯恩针对美国以反恐为名使用无人机对本国公民进行监视和有针对性杀戮的事实，提出这种由中情局执行的秘密杀戮没有经过国会批准，违背了开战的程序正义和合法权威原则，并认为这种杀人行为即使不是战争罪也是不道德的。③ 埃尼马克则聚焦美国中情局使用无人机在巴基斯坦境内发动“定点清除”行动，认为中情局隐秘攻击使“定点清除”行动的审批程序形同虚设，导致美国无人机作战过程中公正使用武力的基础不足，并提出无人机打击的隐秘性增加了其在伦理上的不确定性，破坏了关于约束使用武力的国际规范及美国作为国际规范拥护者的声誉，并强调随之而来的针对官方透明度的需求，即必须根据传统的道德要求以官方机构的权威性公开评估现有的证据，明确“定点清除”行动的合法性和正当性。④ 另一方

① VORSTER N. Killing from a distance: a Christian ethical evaluation of CIA targeted drone killings [J]. The Heythrop Journal, 2015 (5): 836 - 849.

② 周桂银. 国际政治中的外交、战争与伦理 [M]. 南京：南京大学出版社，2018：286 - 287.

③ BYRNE E. Making drones to kill civilians: is it ethical? [J]. Journal of Business Ethics, 2018 (1): 81 - 93.

④ ENEMARK C. Drones over Pakistan: secrecy, ethics, and counterinsurgency [J]. Asian Security, 2011 (3): 218 - 237.

面，人工智能技术的发展导致美国“定点清除”行动合法权威的让渡。人工智能技术的军事应用，推动“战争从实验室打响”从军事家的理想转变为性命攸关的现实。构建智能化条件下的作战实验室，对作战行动进行预先评估，无人机“定点清除”行动日益依赖智能化系统进行战争辅助决策。克诺普斯等人指出：随着人工智能技术在无人机反恐作战中的深入应用，智能化无人作战系统操控主权国家进行开战决策的风险增加，将本应归属于主权国家的开战权利让渡给机器，存在国家意志被机器意志操控的潜在可能。①

最后，探讨无人机“定点清除”行动对正当目的原则的挑战。依据正义战争理论，正当目的原则意味着战争并不能简单地视为政策工具，发动战争是为了谋取更长久的和平、秩序和正义，是为了国际社会的普遍利益，而不是仅仅满足一己私利。当前，学界围绕正当目的原则讨论的一个热点议题，即伴随无人作战系统取代人类作战指挥权力，机器意志可能反向操控国家意志，进而对开战正义的正当目的原则构成挑战。对于无人机“定点清除”行动而言，人工智能技术具备人力无可比拟的运算优势，不仅可以极大地缩短传统战争冗长的作战决策环节，而且可将其运算优势应用于战争结果的模拟推演，使得“战争尚未开始，胜负已在掌握之中”。然而，过于成功的手段优势会导致意图上的盲目。正如沙瑞尔在其著作《无人军队：自主武器与未来战争》中所指出的：“自主武器在速度方面的优势显然是消极的。自主武器会导致加快战争步伐、缩短人类决

① KNOOPS G, JAN A. Legal, political and ethical dimensions of drone warfare under international law: a preliminary survey [J]. International Criminal Law Review, 2012 (4): 697 - 720.

策时间的风险。这将加剧危机的不稳定性。”[①] 一旦战争成为零风险的“机器”，某些军事强国就可以毫不顾忌地使用“以战止战”的方式，先发制人地应对一切战争威胁，使国际安全环境陷入更加动荡不安的局面，这与正当目的的原则在本质上是相背离的。

2. 无人机“定点清除”行动与交战正义

无人机“定点清除”行动作为近年来兴起的一种新的作战方式，其运用过程也需要相应的道德规范进行约束。无人机“定点清除”行动的实际应用，对交战正义的区分原则和相称性原则构成了冲击和挑战。围绕无人化“定点清除”行动对交战正义原则的冲击和挑战，同样引发了学界广泛的探讨。

首先，探讨无人机“定点清除”行动对区分原则的挑战。卡尔皮、伯伦和格拉纳塔指出，无人机在“定点清除”行动中的军事应用带来了战争区分模式的革命性变化。在交战过程中，无人机通过目标识别提取匹配算法规则的有效信息，在一定程度上具备自主做出具体区分判断的能力，从而挑战了人类作为战场区分目标的唯一主体地位，同时也导致无人机“定点清除”行动在遵循区分原则方面面临一系列伦理风险。[②] 对于可能涉及的伦理风险，斯帕罗进一步指出，无人机“定点清除”行动会导致区分原则所规定的退出战斗法则失效。退出战斗法则作为区分原则的重要内容，规定作战人员一旦投降或失去战斗能力，就应当享有退出战斗的权利，纳入禁止伤害的范畴。斯帕罗对无人机系统能够准确辨认人类投降行为深感质疑：通常情况下，人类会采用白旗或举起武器等信号来表示投降，

---

① 沙瑞尔. 无人军队：自主武器与未来战争［M］. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：343.

② KARPPI T，BHLEN M，GRANATA Y. Killer robots as cultural techniques［J］. International Journal of Cultural Studies，2018，21（2）：107－123.

然而，判断投降意图的真伪需要的不仅仅是对物体的识别。本质上，识别投降就是判断对手的意图。[①] 在他看来，依据当前无人机区分能力的发展水平，只能以军服、制式武器等标识物编辑算法来实现识别与区分，难以准确判断人类的意图，难以将需要人类主观意识判断的道德依据转化为算法，结果可能导致算法任意剥夺人的生命，区分原则的退出战斗法则难以真正发挥效力。

此外，还有学者从更深层次探讨，认为美国无人机“定点清除”行动挑战了区分原则存在的道义基础。国际机器人军备控制委员会的创立者阿萨罗指出：区分原则是人类规制战争行为、维护交战正义的重要原则，其作为战争规约的核心精神体现为对人类生命的尊重，而人工智能技术主导下的无人机反恐，本质上是将人类主导区分作战目标的权力转交给人工智能系统，最根本和最突出的道德问题是人的尊严和人权问题。一旦赋予人工智能系统决定人类生死的权力，人工智能系统将以深刻的方式改变整个人类社会的价值观念，进而产生重大伦理风险。这种伦理风险并非主要由基于区分方式的错误所导致，而是因为这种区分方式剥夺了人在决定同类生死上的尊严，其结果必然使杀戮变得毫无人性。

其次，探讨无人机“定点清除”行动对相称性原则的挑战。与区分原则一样，相称性原则同样是交战正义的重要体现，其本质是追求道德成本和道德收益的相称，即根据战争目标和实际情况使用相称的暴力，选择适当的战争方法和手段，避免造成过度的附带伤害，在最短时间内以最小代价结束战争。相称性原则在本质上反映了人类应对控制作战手段的能力，换言之，战争中遵循相称性原则

① 沙瑞尔. 无人军队：自主武器与未来战争［M］. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：292.

集中体现为武器的可控性。当前，学界对于美国无人机“定点清除”行动能否遵循相称性原则方面存在争议。美国佐治亚理工学院的阿金认为，人工智能系统的运算优势可以提升无人机反恐在相称性评估的效率。[①] 这种观点具有一定的合理性，但其存在的问题也是显而易见的。以荷兰阿姆斯特丹大学的布加德为代表的部分学者认为，当前无人机智能系统无法独立进行战术或战略层面的相称性评估。[②] 人工智能技术只能够用来计算简单的、可量化的相称性问题，难以评估复杂问题。沙瑞尔指出：当前无人机的人工智能系统，仅能评估太空、深海、极地等没有平民、相对简单环境下的相称性问题，却难以应对复杂环境下的相称性问题。譬如，人口稠密环境下使用智能化作战平台如何遵循相称性原则？首先需要机器扫描目标周围的平民区域，估计潜在的附带伤害，然后判断是否应该进行攻击，这个过程很难实现自动化。[③] 由此可见，在复杂环境下，权衡军事行动必要性与预期的平民伤亡，涉及极其复杂的道德推理，这是当前智能化作战平台难以企及的。

当前，学界也关注到西方军事强国刻意推行的“大数据战争”和“大数据反恐”。在一些学者看来，其实质上是以数据层面的相称性替代道德价值层面的相称性，将击毙恐怖分子的数量与战争的正义性等量齐观，这种将相称性原则片面数据化的做法存在的伦理风险是显而易见的。钱铖和石斌指出，近年来，美国在利用无人机进行“定点清除”行动时，为了在人口稠密地区完成“定点清除”任

① ARKIN R C. The case for ethical autonomy in unmanned systems [J]. Journal of Military Ethics, 2010, 9 (4): 333.

② BOOGAARD V D. Proportionality and autonomous weapons systems [J]. Journal of International Humanitarian Legal Studies, 2015 (2): 247 - 283.

③ 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：288.

务，往往将攻击目标周围18周岁以上的成年男子默认为恐怖分子，将开车、跑腿的恐怖分子默认为恐怖分子头目。这种片面追求数据相称性的做法，实质上是将复杂的相称性问题粗暴地简单化处理，其结果必然会造成人道主义灾难的进一步扩大化。①

需要注意的是，针对如何应对无人机“定点清除”行动在交战阶段存在的诸多伦理问题，也有部分西方学者乐观地认为，上述伦理问题都可以通过技术创新和技术规制的方式来解决。例如，在无人机“定点清除”行动遵循区分原则方面，斯特劳泽认为，远程虚拟环境下操控无人机面临的区分困境完全可通过虚拟仿真技术手段来规避，通过有针对性的虚拟训练，无人作战模式下对恐怖分子的区分能力不会降低而会增加。② 再如，在无人机反恐遵循相称性原则方面，帕夫利斯切克将无人作战中“规避无效伤害”作为度量反恐打击相称性的标准，提出利用精确制导技术能够规避当前无人机反恐面临的相称性伦理困境。③ 这一观点的实质就是将相称性等同于精准性，借此来验证无人机“定点清除”行动符合相称性规约。“规避无效伤害”标准与西方主流媒体对无人作战平台精确、高效打击能力的宣传相契合，但其在对相称性理解上的局限性也是显而易见的。

上述观点从技术设计维度出发，尝试以技术手段解决无人机“定点清除”行动面临的交战正义挑战，体现了科技专家对无人机“定点清除”行动正义性的人道主义追求，与近代以来兴起的“科

① 钱铖，石斌.“雄蜂”的阴影：美国无人机作战对当代战争伦理的挑战［J］. 世界经济与政治，2013（8）：86－99.

② STRAWSER B. Moral predators: the duty to employ uninhabited aerial vehicles［J］. Journal of Military Ethics, 2010（4）: 353.

③ PAVLISCHEK K. Proportionality in warfare［J］. The New Atlantis, 2010（27）: 23.

学技术和平主义”思潮一脉相承。但是，考虑到战争关乎国家、民族生死存亡的重要性，无人机“定点清除”行动的伦理设计与战争的功利化主题之间必然会产生冲突。需要警惕的是，技术的发展并不能一劳永逸地解决所有的道德问题，反而可能导致少数国家打着“无人化装备使战争更人道”的幌子，刻意夸大技术创新的万能作用，用“零风险”“零伤亡”“低成本”等词汇，为自身打造一面面伦理盾牌，既可以缓解来自国际和国内的道德压力，又能够比以前更方便地使用军事霸权，肆意降低发动无人机反恐作战的道德门槛，同时进一步扩大无人机反恐的作战范围和作战对象范围，使国际安全环境面临更为严峻的挑战。

（三）未来研究展望

虽然围绕无人机反恐的正义性及其相关伦理问题的研究，已经取得了一定的进展，但由于军用人工智能技术仍然处于快速发展的阶段，无人机反恐的作战样式和作战手段在不断丰富和发展，与之相对应的战争伦理理论也有待进一步探讨和完善，未来的研究热点可能在以下几个方面汇聚。

第一，从无人机“定点清除”行动的整个作战流程和技术体系来考察正义性问题。一方面，当前学界对于无人机作战的正义性研究，往往聚焦于无人机“定点清除”式火力打击的正义性考察，却忽视了无人机作战是由情报搜集、分析决策、火力打击等要素构成的复杂作战体系。事实上，各国为获取反恐情报，实施了网络情报战，侵犯了他国网络主权，同时利用大数据预判进行先发制人式反恐，同样引发了一系列伦理问题，应当同样属于美国无人机“定点清除”行动正义性研究的范畴。另一方面，无人机作战样式和作战方法的创新发展，在很大程度上得益于人工智能技术在反恐作战过

程中应用的不断深化，一旦未来强人工智能时代来临，无人机技术体系可能呈现出完全自主化的发展趋势，人类将从指挥控制链条中彻底消失，由此引发的伦理风险也有必要先期进行探索研究。

第二，无人机跨国“定点清除”行动的正当理由问题。特朗普执政期间，美国不仅进一步扩大了无人机反恐的范围，而且以反恐为名对主权国家的军事领导人进行“斩首”行动。美国刺杀伊朗“圣城旅”指挥官苏莱曼尼的事件就是其中的一个典型案例。面对美国利用无人机全球反恐的扩大化趋势，跨国反恐的正当理由始终是一个难以回避的问题。由此可以预见，未来关于无人机反恐伦理中开战正义部分的研究，可能进一步围绕正当理由原则展开，具体探讨的问题包括：哪些无人机“定点清除”行动构成对国家主权的侵犯？哪些侵犯他国主权的跨国“定点清除”行为可认定为侵略并可实施反击？在面临他国无人机侵犯国家主权的严峻风险时，主权国家如何在遵循区分原则和相称性原则的前提下进行自我防卫，是否可以采取“跨域攻击”的方式进行反击（即反击的维度不局限于本国领空范围）？

第三，无人机“定点清除”行动的战后正义讨论。当前围绕无人机“定点清除”行动的正义性研究，主要集中在开战正义和交战正义两个环节，事实上，战后正义也是正义战争理论体系的重要组成部分。传统的战后正义的实现方式，包括战后重建、惩治战犯以及战争赔款等。由于无人机“定点清除”行动不同于旷日持久的传统战争，且主要采取远程精确制导的方式进行，其物理毁伤程度与常规战争的“地毯式轰炸”“饱和攻击”不可同日而语，因此，围绕无人机“定点清除”行动的战后正义讨论，目前尚未引起学术界的高度重视。然而，随着无人机“定点清除”行动在全球范围内的持续深化，误伤平民的责任追究、反恐交战范围内民众的心理创伤、

战后秩序恢复和反恐作战治标不治本等一系列问题日益凸显，可以预见，未来围绕无人机“定点清除”行动的战后正义问题，将有望成为新的研究热点。

此外，还需要注意无人机“定点清除”行动的伦理规制研究。当前围绕无人机“定点清除”行动的伦理对策研究，主要分为两个流派。一派是传统的正义战争理论研究者，其研究思路主要是“问题倒逼”，即修正传统的正义战争伦理体系，以规制无人机“定点清除”行动带来的新问题、新挑战；另一派是具有人工智能领域专业背景的科学家和工程技术专家，其研究思路主要是“技术设计”，即通过预先对无人作战系统的伦理设计，使其匹配现有的战争伦理规约。两派研究长期以来相互分离，根本问题在于前者主要关注“定点清除”行动应用层面的伦理问题，后者则聚焦作战技术、武器装备平台设计与研发层面的伦理问题。随着无人化武器研发与作战应用之间联系与互动日益紧密，未来上述两个流派可能会出现合流的趋势：传统伦理流派可能更加关注不断演进的无人作战伦理规约对武器装备设计研发产生的动态影响；技术流派可能会关注无人化武器装备在“定点清除”行动应用过程中不断创新发展而引发新的伦理问题。

三、智能识别方式下火力打击的区分困境

火力打击是从物理上摧毁目标的关键环节，是无人机“定点清除”行动流程中最吸引外界注目的一环，也是引发伦理道德争议最多的一环。美军通过远程遥控方式，指挥无人机在万里之遥的他国领空“定点清除”恐怖分子，成为当前无人机反恐作战的主要火力打击模式。各种由人类远程操控、具备目标自动识别和攻击能力的

智能化反恐装备扮演了终结者的角色，如“察打一体”的“捕食者”无人机和“收割者”无人机等，对己方的“零伤亡”与对敌方的“定点清除”式精确打击，在一定程度上契合了正义战争的道德准则。但是，这种看似低风险甚至无风险的无人机“定点清除”行动方式一旦被滥用，就会给现代战争打上厚重的“非人化”烙印，不仅对国家安全构成重大威胁，而且给战争伦理体系提出新的道德挑战。

锁定目标是火力打击的前提条件。作为人道主义精神的体现，区分原则一直是正义战争理论的重要内容。区分原则指的是“非战斗人员的生命和财产以及某些种类的民用目标不应该被作为军事暴力的目标”，其核心在于非军事人员与目标的豁免权。[①] 作为一种故老相传的观念，战争就是战斗员之间的战争。早在远古时期的战争，就形成了区分军民的朴素规约。伴随智能化科技手段的创新发展，无人机“定点清除”行动的火力打击环节是在虚拟环境下远程操控进行的。无人机操控员可以利用智能识别设备，实现战场形势的远程实时感知和“发现即摧毁”式“秒杀”，在一定程度上弥补了人类视距上的局限。然而，在虚拟化的远程火力打击环境下，无人机“定点清除”行动在火力打击环节同样面临锁定目标的区分困境。下面以“定点清除”行动为例，分析无人机作战在火力打击环节面临的伦理挑战与风险。

（一）虚拟的作战环境

古往今来的人类战争一直在现实的作战环境中展开。在传统作

① 考彼尔特斯，福臣，时殷弘. 战争的道德制约：冷战后局部战争的哲学思考［M］. 邹琳，戴锋宁，等译. 北京：法律出版社，2003：155.

程中应用的不断深化，一旦未来强人工智能时代来临，无人机技术体系可能呈现出完全自主化的发展趋势，人类将从指挥控制链条中彻底消失，由此引发的伦理风险也有必要先期进行探索研究。

第二，无人机跨国“定点清除”行动的正当理由问题。特朗普执政期间，美国不仅进一步扩大了无人机反恐的范围，而且以反恐为名对主权国家的军事领导人进行“斩首”行动。美国刺杀伊朗“圣城旅”指挥官苏莱曼尼的事件就是其中的一个典型案例。面对美国利用无人机全球反恐的扩大化趋势，跨国反恐的正当理由始终是一个难以回避的问题。由此可以预见，未来关于无人机反恐伦理中开战正义部分的研究，可能进一步围绕正当理由原则展开，具体探讨的问题包括：哪些无人机“定点清除”行动构成对国家主权的侵犯？哪些侵犯他国主权的跨国“定点清除”行为可认定为侵略并可实施反击？在面临他国无人机侵犯国家主权的严峻风险时，主权国家如何在遵循区分原则和相称性原则的前提下进行自我防卫，是否可以采取“跨域攻击”的方式进行反击（即反击的维度不局限于本国领空范围）？

第三，无人机“定点清除”行动的战后正义讨论。当前围绕无人机“定点清除”行动的正义性研究，主要集中在开战正义和交战正义两个环节，事实上，战后正义也是正义战争理论体系的重要组成部分。传统的战后正义的实现方式，包括战后重建、惩治战犯以及战争赔款等。由于无人机“定点清除”行动不同于旷日持久的传统战争，且主要采取远程精确制导的方式进行，其物理毁伤程度与常规战争的“地毯式轰炸”“饱和攻击”不可同日而语，因此，围绕无人机“定点清除”行动的战后正义讨论，目前尚未引起学术界的高度重视。然而，随着无人机“定点清除”行动在全球范围内的持续深化，误伤平民的责任追究、反恐交战范围内民众的心理创伤、

战后秩序恢复和反恐作战治标不治本等一系列问题日益凸显，可以预见，未来围绕无人机“定点清除”行动的战后正义问题，将有望成为新的研究热点。

此外，还需要注意无人机“定点清除”行动的伦理规制研究。当前围绕无人机“定点清除”行动的伦理对策研究，主要分为两个流派。一派是传统的正义战争理论研究者，其研究思路主要是“问题倒逼”，即修正传统的正义战争伦理体系，以规制无人机“定点清除”行动带来的新问题、新挑战；另一派是具有人工智能领域专业背景的科学家和工程技术专家，其研究思路主要是“技术设计”，即通过预先对无人作战系统的伦理设计，使其匹配现有的战争伦理规约。两派研究长期以来相互分离，根本问题在于前者主要关注“定点清除”行动应用层面的伦理问题，后者则聚焦作战技术、武器装备平台设计与研发层面的伦理问题。随着无人化武器研发与作战应用之间联系与互动日益紧密，未来上述两个流派可能会出现合流的趋势：传统伦理流派可能更加关注不断演进的无人作战伦理规约对武器装备设计研发产生的动态影响；技术流派可能会关注无人化武器装备在“定点清除”行动应用过程中不断创新发展而引发新的伦理问题。

三、智能识别方式下火力打击的区分困境

火力打击是从物理上摧毁目标的关键环节，是无人机“定点清除”行动流程中最吸引外界注目的一环，也是引发伦理道德争议最多的一环。美军通过远程遥控方式，指挥无人机在万里之遥的他国领空“定点清除”恐怖分子，成为当前无人机反恐作战的主要火力打击模式。各种由人类远程操控、具备目标自动识别和攻击能力的

智能化反恐装备扮演了终结者的角色，如“察打一体”的“捕食者”无人机和“收割者”无人机等，对己方的“零伤亡”与对敌方的“定点清除”式精确打击，在一定程度上契合了正义战争的道德准则。但是，这种看似低风险甚至无风险的无人机“定点清除”行动方式一旦被滥用，就会给现代战争打上厚重的“非人化”烙印，不仅对国家安全构成重大威胁，而且给战争伦理体系提出新的道德挑战。

锁定目标是火力打击的前提条件。作为人道主义精神的体现，区分原则一直是正义战争理论的重要内容。区分原则指的是“非战斗人员的生命和财产以及某些种类的民用目标不应该被作为军事暴力的目标”，其核心在于非军事人员与目标的豁免权。[①] 作为一种故老相传的观念，战争就是战斗员之间的战争。早在远古时期的战争，就形成了区分军民的朴素规约。伴随智能化科技手段的创新发展，无人机“定点清除”行动的火力打击环节是在虚拟环境下远程操控进行的。无人机操控员可以利用智能识别设备，实现战场形势的远程实时感知和“发现即摧毁”式“秒杀”，在一定程度上弥补了人类视距上的局限。然而，在虚拟化的远程火力打击环境下，无人机“定点清除”行动在火力打击环节同样面临锁定目标的区分困境。下面以“定点清除”行动为例，分析无人机作战在火力打击环节面临的伦理挑战与风险。

（一）虚拟的作战环境

古往今来的人类战争一直在现实的作战环境中展开。在传统作

① 考彼尔特斯，福臣，时殷弘. 战争的道德制约：冷战后局部战争的哲学思考［M］. 邹琳，戴锋宁，等译. 北京：法律出版社，2003：155.

战环境中，战士在与对手面对面的厮杀搏斗中能够清楚地看到对手的身体，在白刃战和肉搏战中与对手保持充分的身体接触，因此可以通过视觉、听觉、嗅觉等感官来感受对手带来的压力。在传统作战环境中，战士也必须经常与战友保持密切联系。这种传统的作战环境和作战模式不仅具有军事上的影响，而且对作战人员具有重要的认知和道德影响。比如，每一个身处其中的战士都清楚地知道，他面前的对手是一个与他一样鲜活的人，也同样在努力渴望取得战争的胜利。在交战过程中，他会在第一时间收到来自对手的各种反馈，他能够对战场中对手的处境感同身受。概而言之，在传统战争模式下，战士全身心地投入战斗，高度集中地沉浸在战场中，由于精神与身体高度统一，因此所做的一切都会遇到具体的身体上的阻力。

无人机“定点清除”行动的一个重要特征，就是利用远程虚拟识别环境，使人类作战环境由自然空间向虚拟空间转换，无人机可以通过卫星通信在遥远的地方执行作战任务，作战人员可以通过卫星实时传输的视频感知战场态势，从而远离危险，实现作战人员的零伤亡，降低交战成本。然而，虚拟空间也使区分原则的执行面临新的挑战。虚拟空间对区分原则的挑战，主要体现为可能助长操控者的游戏心态，进而造成附带伤害的扩大化，具体体现在两个方面。

一方面是虚拟环境造成的游戏心态。无人作战特殊的作战方式，给予操控人员一种特殊的疏离感。在虚拟战场环境下，操控者与作战目标的距离不仅没有缩短，反而更加遥远了。他们处于距战场万里之遥的控制中心，坐在远离战场的软垫座椅上。无人机操控员视听感官与真实战场相互隔绝，依据卫星传递的实时画面，用操控杆和油门踏板来遥控战争，无人机“定点清除”行动完全可以通过计算机屏幕和远程声频反馈来实施，由此可能导致无人机操控者的行

为逐渐与其所产生的后果割裂开来。当杀戮得以在虚拟世界中进行，区分原则对无人机操控员的约束力也会随之减弱，致使他们更加轻易地做出判断，对眼前的镜像产生视频游戏的错觉，以游戏心态来看待杀人的风险。一旦操控员以近乎游戏的心态对待战争，区分原则就难以严格执行。

另一方面是情境切换带来的心理反差。不同于传统飞行员，无人机操控员每天都可以正常上下班，因此需要在无情的杀人任务与温情的家庭生活中往复切换，这对他们的心理承受能力是一个挑战。在执行“定点清除”任务时，他们依靠实时监控影像锁定目标，画面中恐怖分子的生活场景可能会勾起某种熟悉感，造成工作情境与生活情境的紊乱，这进一步增加了锁定和打击目标的难度。在英国电影《天空之眼》中，执行任务的无人机操控员在视频监控中看到了一个小女孩和她家庭的温馨场景，导致在随后的“定点清除”行动中，他不愿意冒着牺牲无辜女孩生命的风险去击毙恐怖分子。

在虚拟环境下，无人机操控员还面临“机会目标”的心理压力。在某些反恐作战中，操控员需要面对高速流动的战场环境，无人机始终以挂载武器的形式执行巡航任务，在发现紧急情况时打击地面上的“机会目标”。在有限的时间跨度内，面对高速流动的战场，无人机操控者的“时间压力”进一步受到“游戏心态”的影响，更倾向于做出把握“机会目标”的决策。虽然美军在对外宣称中显示出对区分原则的尊重，奥巴马总统的反恐顾问布瑞南曾经宣称“美国政府在使用无人机技术时是非常慎重的”，并指出无人作战需要满足四个条件：一是攻击者正在面临迫切的威胁；二是攻击者确定实施抓捕是不可行的；三是攻击者承认在海外单方面的实施行动前，需要检查作战目标的重要性；四是攻击者对确定目标身份及避免平民

伤亡有高度自信。[①] 但是，上述四个条件仍然为潜在的附带伤害留下了余地。迫切威胁的评估、目标重要性的衡量和避免平民伤亡的信心，都存在极大的不确定性。受游戏心态鼓舞的主观意志进一步加大了对“机会目标”做出正确判断的难度。

对于虚拟识别环境下的游戏心态，部分学者也提出不同意见，认为“无人作战带来的认知困境挑战操控员区分能力”的论点是难以成立的。在斯特劳瑟看来，虚拟环境下的游戏心态完全可通过仿真技术手段来规避。利用虚拟现实技术，无人作战模式下操控员的区分能力不会降低，而会提高。由于无人作战对于远程操控者而言毫无风险，操控员不会焦虑不安，他们可以用更长的时间区分作战目标，因而可能做出更明智的决定。[②] 该观点虽然从技术层面为规避认知困境提供了一种可能，却忽略了当前无人作战的现实境况。事实上，虚拟识别环境对区分原则的挑战，在深层次上反映了无人机给予战争的“非人化”烙印。如果执行“定点清除”任务对于作战人员而言毫无风险，作战人员以人性角度感知理解战争的能力就可能会逐渐丧失，从而对战争的道义基础造成根本性的冲击。

（二）“盲人摸象”式的特征攻击

目前，在无人机“定点清除”行动中，美军在阿富汗、巴基斯坦、也门等地推行极富争议的特征攻击，这对区分原则构成了进一步的冲击。在真实战场环境下的常规战争中，区分军民目标主要依靠人的感官经验，因为两者在装备和服饰上存在显著的差别。为应

① DAVIS L E, MCNERNEY M J, CHOW J S, et al. Armed and dangerous? UAVs and U.S. security [R]. RAND, 2014.

② STRAWSER B. Moral predators: the duty to employ uninhabited aerial vehicles [J]. Journal of Military Ethics, 2010 (4): 353.

对北约国家在军事技术上的绝对优势，阿富汗等地区的恐怖分子采取混杂于平民之中的策略，以免成为无人机锁定的靶子。面对人类感官经验难以发挥效用、战斗人员与非战斗人员的区分变得困难等境况，在真实身份尚未明确的情况下，美军对满足某些预设恐怖分子特征（如携带刀具或其他管制武器）的目标实施打击。[①] 依据这一方式，美国中央情报局可以操控无人机在未经确认身份和未经国会授权的情况下，对战区范围以外、具有与敌对分子相匹配的标识或特征的目标实施“定点清除”，只要这个人被技术分析认为是“基地”组织或其他恐怖组织的“高级指挥官”，即便是对于本国公民也不例外。

特征攻击模式是在军民混杂战场环境下执行区分原则的一种尝试，这种根植于虚拟世界识别区分技术的火力打击手段，虽然具有其先进性的一面，但并不足以替代现实世界人类感官的验证，其依托信息技术规避传统区分困境的意图是毋庸置疑的，其片面依赖特征匹配的局限性也是显而易见的，必然是与区分原则的精神背道而驰的。美国康奈尔大学的克雷普斯指出，特征攻击对于目标行为方式的偏好，并不能将其与恐怖分子的关联性等量齐观。[②] 在“定点清除”行动中，用特征匹配的区分方式取代传统的区分方式，还引发了关于无人机附带伤害的争议。正如拜门在《外交事务》杂志上所指出的：“所谓的‘特征打击’，目标不是具体的个人，而是活动的群体。这一方式就更难区分军事人员和平民了。”[③] 区分原则禁止

① STRAWSER B. Moral predators: the duty to employ uninhabited aerial vehicles [J]. Journal of Military Ethics, 2010 (4): 353.

② KREPS S. Flying under the radar: a study of public attitudes towards unmanned aerial vehicles [J]. Research and Politics, 2014 (2): 2.

③ BYMAN D. Why drones work, the case for Washington's weapon of choice [J]. Foreign Affairs, 2013 (4): 32 - 43.

攻击平民、战俘以及丧失作战能力的士兵，然而，这一原则未能完全禁止因攻击合法目标而造成的非战斗人员伤亡，这就给利用特征攻击造成的附带伤害预留了空间。一方面，美国政府否认无人作战的附带伤害，认为平民伤亡是极其罕见的，甚至在“个位数”，其目的在于强化无人作战“精确、无风险”的形象；另一方面，作为非营利性的民间机构，英国新闻调查局提出了学界认可的数据，在2004至2020年间，无人机在巴基斯坦至少发起430次攻击，造成平民死亡人数多达969人。[①] 美国政府惯常将所有死伤的成年男性默认为武装分子，实质上是对区分原则和附带伤害概念的扭曲。

（三）“非人化”的技治主义倾向

无人机“定点清除”行动中使用特征攻击的火力打击方式，表面上是由虚拟作战环境下的认知困境所导致的，在深层面则根源于识别锁定目标过程中“非人化”的技治主义倾向。在军民混杂、高速流动的反恐战场，依靠服饰、武器等传统的标识性特征，难以确认恐怖分子的身份。有鉴于此，无人机作战的目标判定，更多地借助人工智能技术手段，人工智能技术主导的区分方法虽然在一定程度上提升了辨识目标的效率，但是也导致了在无人机火力打击上的技治主义倾向。片面强调技术手段和数量方法，造成无人机反恐在火力打击环节面临诸多道德问题。一是对杀人的麻木心态。技治主义者会陷入机械决定论或因果一义论的思维模式，一方面把世界视作一架庞大的机器，另一方面轻视甚至无视人与社会之间问题的多变性与复杂性，造成重物不重人、见物不见人的倾向，对于剥夺他

① 张煜，杜雁芸. 人工智能军事化发展态势及其安全影响［J］. 外交评论，2022(3)：120.

人生命的行为产生麻木心态。二是以技术数据"绑架"舆论。在电影《天空之眼》中，北约联合反恐指挥官为了实现打击恐怖分子的目的，不惜命令下属人员刻意篡改火力打击效果评估数据，从而导致无辜女孩的丧生。事实上，在真实反恐作战中也存在类似的问题。由于反恐火力打击依赖精细的数据处理和技术分析，因此，无人机作战相关数据的控制权掌握在技术专家和管理人员手中，可能造成民众难以了解无人作战附带伤害的真实情况。美军官方公布数据与民间调查数据之间的巨大差距，说明存在政府、军方和技术专家"绑架"舆论的可能性。

鉴于无人机"定点清除"行动在火力打击环节的区分方式存在上述弊端，在继续使用必要的技术区分手段的同时，还必须重视感官的直接体验。感官判断虽然具有一定程度的主观性，但同样是区分军民不可或缺的重要凭据。要在目标区分上实现技术与感官的结合，就必须削弱当前的技治主义倾向，打破区分权集中在政府、军方和技术专家手中的现状，实现权力的分散化，将土生土长的当地民众作为参与者，介入目标搜索、识别和判断的全部过程，充分尊重其感官在目标识别中的参考价值，这也许是突破当前无人作战区分瓶颈的一条可行路径。

四、"平台无人"模式下的武力控制困境

无人化武器平台的出现，原本可视为武力可控作战的一个典型范例。以执行"定点清除"任务的无人机为例，"定点清除"式的无人作战至少在四个方面实现了可控：一是时间可控，力求速战速决；二是空间可控，按既定路线巡航，限定作战地域；三是目标可控，依托高精度制导炸弹，选择性杀伤有限目标；四是结局可控，

实施不对称打击，确保达成预期目标。在西方某些政治家的宣传中，无人作战是以精确控制来占据道义的制高点。然而，问题就在于，在无人操控环境下，西方军事强国凭借“机器 VS 人”的不对称打击，轻易地实现了“零风险、零伤亡、低成本、非接触”，作战目标层面的火力精确可控往往以作战手段层面的不可控作为代价。在围绕无人机“定点清除”行动的相关道德争议中，一个突出的问题是：无人机是不是一种可控的武力打击平台？正义战争理论要求，应当避免超出预期军事目的所需的过度杀伤，就是指作战手段的运用必须可控。在实现武力打击的可控性方面，无人机在“定点清除”行动中的应用同样面临诸多挑战。

一是战略层面武力运用的不可控。在人类自觉的对象性活动中，手段和目的是两个密切关联的要素。人们总是以一定的手段来实现既定的目的，然而，手段的过于成功往往会导致目的的迷失。对于无人作战而言，在控制中心用操控杆发射导弹，就可以决定万里之外的土地上人民的生死，这对于任何一个国家而言都是一种过度的权力，权力欲的膨胀必然会导致武力使用的失控。美军使用无人机执行“定点清除”行动，犹如上瘾般愈演愈烈，正是无人机在战略层面不可控的鲜活反映。

在无人机“定点清除”行动的作战环境下，无人机的武力控制困境，集中体现在力量对抗的不均衡导致战略层面的不可控，由此也导致对开战正义原则的冲击。开战正义的道义标准，为发动战争的国家确立了道德门槛。事实上，开战正义得到认可的基础，除了道义上的因素，还在于交战双方军事力量对抗的均衡性。一方控制使用武力的手段和方式，是为了避免另一方的过激反应导致战争的升级。在战略博弈中，如果双方的军事实力呈现基本对等的局面，博弈双方就会审慎思考战争的残酷性和取得战争胜利的惨重代价，

从而将军事行动控制在一定烈度内，避免冲突的升级。冷战时期美苏之间的长期对峙，即是军事战略博弈长期维持均势的一个缩影。然而，自海湾战争以来，随着信息技术在军事领域的广泛应用，西方军事强国日益依靠不对称的技术优势来打破均衡性的约束，将战争的结果从无法预知转变为可以预知。“捕食者”等无人化作战平台的出现，进一步强化了美国等北约国家的不对称武力打击优势。较之常规武器与常规作战方式，无人机摆脱了人类体能的桎梏，能够在无人操作的环境下实现全球快速打击与隐蔽攻击，因而降低了发动战争的成本和门槛，进一步挑战了维系相称性原则的均衡性。情况正如澳大利亚麦考瑞大学的加里奥特所述，技术上的极端不均衡，导致技术后进国能够给予技术先进国的伤害变得微不足道，相称性原则的执行也就变得难以为继。[①] 在巴基斯坦、也门等地，由无人机实施的“机器 VS 人”的不对称打击，已经使发动战争变得毫无风险，均衡性的丧失必然导致无人机的滥用，进而挑战其作为智能化作战手段的可控性。一旦战争的代价降低甚至消失，军事强国就可以无视开战正义的道德标准，以反恐为名行强权之实。

二是战术层面武力运用的不可控。无人机反恐“定点清除”行动的不可控性还体现在战术层面，具体而言，就是对交战正义中相称性原则的挑战。经典战争伦理体系不仅涵盖发动战争之前的开战正义，还包括交战过程中的道德标准，主要体现为区分原则和相称性原则。作为对区分原则的必要补充，相称性原则在交战正义的伦理规约中同样占据举足轻重的地位。依据相称性原则，军事目标的预期价值与军事打击的潜在伤害之间必须维系一种平衡。区分原则

① GALLIOTT J C. Uninhabited aerial vehicles and the asymmetry objection: a response to strawser [J]. Journal of Military Ethics, 2012 (1): 63.

强调的是对非军事目标的豁免权，相称性原则关注的是对作战手段的控制。即使面对区分之后的合法攻击目标，也不能在程度或数量上造成超过正当限度的损伤，相称性原则的主旨在于抑制武力的滥用。

无人机的不对称打击优势，造成在战术层面武力运用的不可控。作为远程精确打击手段，无人机的零风险与绝对优势，使得美军在无人机的战术运用中不再谨慎小心。斯坦福国际人权中心的调查报告指出，美军无人机的打击范围日益向葬礼、聚会和社区会议等人口密集地域拓展。美国《新闻周刊》记者也撰文批判名为“两次打击”的无人机战法，认为这种对同一作战地域实施相对连续的两次打击以确保目标死亡的做法，会加剧人道主义救助者和其他无辜者的伤亡。[①] 把一种精确打击手段转变为饱和攻击的手段来使用，无人机在战术层面的滥用可窥一斑。

无人机反恐“定点清除”行动在战术层面的滥用，在深层次体现为可控标准界定的模糊。用何种标准来衡量军事目标价值与军事打击伤害之间的比例，以及如何界定作战手段是否可控？战争伦理学界并没有一个公认的标准。目前，部分西方学者将“规避无效伤害”作为度量可控性的标准。[②] 其实质就是将可控性等同于精准性，以此来验证无人机作战符合交战正义的规约，这一标准与西方主流媒体对无人机精确、高效打击能力的宣传相契合，但其在对可控性理解上的局限性也是显而易见的。即便不考虑官方关于无人机打击数据的可信度，将精确性作为衡量控制能力的做法也存在诸多弊端，

① GREENWALD G. U. S. drones targeting rescuers and mourners [EB/OL]. (2012 - 02 - 05) [2021 - 02 - 23]. https://www.salon.com/2012/02/05/u_s_drones_targeting_rescuers_and_mourners/.

② PAVLISCHEK K. Proportionality in warfare [J]. The New Atlantis, 2010 (1): 23.

因为精确性并不能区分打击目标的重要程度，所以会导致对底层恐怖分子更为频繁的攻击。

在“定点清除”行动中对无人机的过度依赖，必然造成对交战正义中相称性原则的冲击。情况正如沃尔泽所言：“较之其他杀伤形式，无人机容易太多了，以至于让我们产生不安。它是既诱人又危险的技术手段，使我们的敌人比以往更脆弱，而我们的士兵没有任何风险。……但技术太好，使用的标准就会逐步放松。”[①] 依据英国新闻调查局的数据统计，在美国军队和中央情报局动用无人机击毙的目标中，仅有不到2%的高价值目标（恐怖组织头目）。对于剩余98%的低价值目标（底层恐怖分子）动用“地狱火”导弹是否妥当？这引发了美国政府内部高层人士对于反恐效费比的反思。更为值得关注的是，奥巴马执政期间，美国政府采取了一种颇具争议的划分方法，即将作战区域内所有达到服役年龄的男性预设为攻击目标。这种不加区分的攻击方式，实质上为无人机进一步的滥用大开方便之门。

无人机“定点清除”行动在战术和战略层面产生的武力控制困境，根源在于短期相称性与长效相称性原则的冲突。作为对战争规约的功利主义诠释，英国伦理学家西季威克对相称性原则的定义是：在战斗中不允许造成任何实质性无助于胜利目标的危害。[②] 相称性原则作为交战正义的重要准则，本身即包含两个方面的含义：一是当下的相称性，即作战的效用必须与其造成的直接伤害相称；二是长效的相称性，即作战的效用还必须与对人类长期利益的损害相称。

少数西方媒体刻意强调无人机作战满足“当下的相称性”的一

① WALZER M. Just & unjust targeted killing & drone warfare [J]. Daedalus, 2016 (4): 12.

② SIDGWICK H. Elements of politics [M]. Adamant Media Corporation, 2000: 254.

面，通过发动舆论工具，大肆宣扬无人机击毙恐怖组织头目、清除武装分子避难所、重创“基地”组织和相关反美军事集团，彰显其所取得战果与造成损伤之间的相称性。然而，西方国家政府及其主流媒体对无人机反恐与长效相称性背离之处却三缄其口。无人机“定点清除”行动作为一种武力手段运用，其与长效相称性的冲突，在对敌和对己两个方面均有体现。“以暴制暴”的方式并不是反恐的治本之策，实施肉体摧毁也不能从源头上清除产生恐怖主义的根源，相反，还会加剧当地民众的仇恨情绪和动荡局势。事实上，“9·11”事件之后，美国在全球展开的反恐攻势频繁陷入“愈反愈恐”的尴尬境地，与其对长效相称性的忽视密切关联，这也是造成美国反恐战略性失误的深层次原因。鉴于“以暴制暴”无法从源头上清除恐怖主义产生的根源，因此改变刻意追求反恐杀伤效率的做法，适度减少无人机反恐的使用频率，审慎对待潜在的打击目标，注重无人作战的长效相称性，才是缓解当地民众的仇恨情绪和动荡局势，在长远时间内应对恐怖主义泛滥的治本之策。

五、“机器 VS 人”攻击模式下的责任困境

经典正义战争理论由开战正义、交战正义和战后正义三个部分内容构成。战争责任是对战后正义的重要检验，在沃尔泽看来，只要我们准确地指出责任者，或者至少我们的责任分配和道德判断符合战争的实际经验，对战争的所有痛苦保持敏感，就会极大地增强正义论的力量。在经典正义战争论的理论体系中，责任主体聚焦于君主、官员、军人等特定人群。在无人机“定点清除”行动中，实现战后正义面临责任困境的挑战。

无人机“定点清除”行动，必然存在由谁来为火力打击后果负

责任的问题。随着无人机系统智能化水平和自治能力的不断提高，它由原来纯粹的作战手段转变为逐渐具备战争主体部分功能。无人作战条件下人机集成共同扮演火力打击的主体角色，冲击了正义战争理论对于责任主体与责任分配的固有认识，引发了一系列的争议。

（一）责任主体争议

在“机器 VS 人”攻击模式下，无人机能否成为更为人道的责任主体？随着人工智能技术的不断发展，以“捕食者”为代表的部分无人作战平台已初步具备了智能化的某些特征，由此引发了关于战争责任主体的新的争议。即相对于传统战争中人类作为唯一的责任主体，智能化的无人作战平台能否更好地承担火力打击之后的一系列战争责任，进而将未来战争引向更为人道的方向？对于这一问题，无人机技术的支持者采取了乐观的态度。阿金指出，在确保交战正义性上，较之有人作战平台，智能化的无人作战平台具有以下六个优势：无须考虑自身安全；具备超人的战场观察能力；不受主观情绪左右；不受习惯模式影响；更快的信息处理速度；独立、客观地监测战场道德行为。[①] 基于以上优势，阿金认为在执行人道主义原则方面，无人作战平台会比人类表现得更好。

然而，一旦无人机取代人类成为未来战争攻击行动的决策者，人道主义原则的执行可能面临诸多挑战。一是鉴于未来人工智能发展前景的不确定性，无人机做出符合人道规约的作战决策尚面临难以克服的技术瓶颈。二是无人机自治系统存在不稳定性和巨大的风险性。控制系统故障、电子信号干扰、黑客网络攻击和其他战场上

① ARKIN R C. The case for ethical autonomy in unmanned systems [J]. Journal of Military Ethics, 2010, 9 (4): 333.

可能发生的意外情况，都会影响其执行符合人道主义规约的决策，甚至造成战场杀人机器的失控。三是无人机取代人类进行作战决策，意味着它由战争工具转变为战争主体，而人道主义原则作为约束人类行为的伦理规约，难以适用于无人机相互厮杀的战争。

（二）责任分配困境

人类与智能化的无人机之间如何进行责任分配？无人机“定点清除”行动对于责任分配的挑战，首先体现为主体多元化加大了责任追究的难度。在常规的战争模式下，战争机器由权责分明的军事部门和政府机构来推动，因此，对于战犯的指认与辨识在操作层面并非难事。正如沃尔泽所言：“军人和政治家就不得不作出选择，有时是道德的选择。既然有人选择，就必定可能把他们指出来加以称赞和谴责。如果有可以辨认的战争罪行，就必定有可以辨认的罪犯。”① 然而，对于因无人机作战引发的非正义军事行为，其责任的追究可能涉及操控者、军火商、程序员、采购官、战地指挥官、相关维护人员乃至无人机本身，涵盖研制、生产、装备与应用的整个过程，其追责难度较常规战争进一步提升。此外，鉴于无人机与操控员之间的匹配问题，为了提升人机结合的效率，美军近年来开始尝试利用脑机接口技术，将大脑信号输出直接用于控制外部武器系统，其中就包括无人作战系统。未来，实现人机一体化的半机械人可能成为无人机的实际操控者，一旦人脑与无人机的自治系统并驾齐驱，战争责任的分配必将陷入更大的困境。

此外，无人机导致的责任分配困境，还体现为如何应对“责任

① 沃尔泽．正义与非正义战争：通过历史实例的道德论证［M］．任辉献，译．南京：江苏人民出版社，2008：319.

转嫁”的问题。无人自治系统替代人脑进行作战决策，虽然可以规避人类情感弱点在战时可能引起的混乱与失控，摒除怜悯、懦弱、愤怒、仇恨等感性因素对作战决策的影响，但是无人自治系统的可控性同样存在巨大的隐患。英国人工智能和机器人专家沙基撰文指出：人为错误、人机交互失败、系统失灵、通信未升级、软件编码错误、敌人网络攻击或渗透供应链，干扰、电子欺骗、敌人的其他行为或是战场上的意外情况等任何错误的发生，都可能造成这类“杀人机器人”在战场中失控，从而将战争引入完全不可预知的结局。针对无人作战系统的安全漏洞，潜在的敌对国家或组织可以通过干扰或欺骗等技术手段，破坏无人机的数据通信链路，劫持并操控其执行有悖人道主义精神的军事行动，同时将相应的战争责任转嫁给无人机的所有国。利用人机的高度一体化，军方和政府可以把人为的责任转嫁给无人机，以逃避战争罪责。2011 年，伊朗的电子战专家通过重构 GPS 坐标成功俘获美国 RQ－170“哨兵”无人机，证明实施“责任转嫁”在技术上是完全可行的。一旦被俘获的无人机用于执行军事任务，鉴于当前网络电磁攻击的溯源难度，真正厘清无人攻击的责任主体是难以实现的。

六、无人作战的人机融合：挑战与出路

自“9·11”事件以来，伴随美国“反恐战争”在全球如火如荼地展开，智能化无人作战平台成为美军“定点清除”恐怖分子的主要手段。与之相对应，无人作战的人机结合问题也日益凸显：自然人的生理和心理素质如何匹配高度智能化的机器人？人与机器两者之间如何实现深度融合，进而最大程度发挥作战系统的全部潜能？这些成为各国军事家共同关注的问题。美国空军前参谋长威尔士指

出："莱特兄弟飞机的时代已经过去，我们正处在第二个阶段。在接下来的二十到三十年，无人机发展将是爆炸性的。可以激动地看到，我们的空军处在且必须处在领导地位，因为我们知道使用它们的最佳方式。创新就是一切。"①

（一）人机融合的挑战与风险

1. 军人生理和心理能力素质的挑战

军队战斗力的发挥有赖于人与武器的结合，在理想状态下，人与武器都能够充分发挥各自的优势，从而发挥"人－机"作战系统的最佳效能。智能化作战方式的广泛应用，丰富了现代战争的攻击方式与作战手段，同时使传统的人装结合模式面临严峻挑战，其独特的作战方式给操控员带来一种疏离感。具体而言，他们坐在远离战场的软垫座椅上，用操控杆和油门踏板操控战斗。在远离真实战场的虚拟环境下，如何实现人与无人机之间的紧密结合，已经成为影响美军无人机反恐作战效率提升的一个难题。人类日益成为无人作战系统中最弱的一环。当前方兴未艾的无人作战，当武器成为具有人工智能系统的机器人，人类能力素质层面固有的生理、心理劣势，就成为制约人与武器有机结合的短板弱项。

一方面，就生理层面而言，在大数据时代，人类与生俱来的生理机能愈发不能适应对海量数据进行快速有效处理的现实需求。执行"定点清除"行动的无人机操控员在战场感知能力方面难以适应无人作战系统高速运转的节奏。回顾历史，从1997年IBM公司研发的"深蓝"计算机击败国际象棋冠军卡斯帕洛夫，到2017年

① HOAGLAND B T. Manning the next Unmanned Air Force: developing RPA pilots of the future [R]. Brookings, 2013: 1.

DeepMind 公司开发的阿尔法围棋（AlphaGo）击败职业围棋世界冠军柯洁，借助机器学习与深度学习技术，人工智能系统在某些方面具备人类难以匹敌的数据处理能力。在军事智能化浪潮的冲击下，人类作为人工智能武器系统的关键操作员，自身的反应速度越来越难以匹配智能化战争模式下自主武器系统所需的运转速度，这直接威胁到人类作为战争主体的参与程度。较之传统飞行员，无人机飞行员需要具有更强的战场感知能力。在虚拟的战场环境下，他们必须具有更高的专注度，及时通过指挥中心屏幕上的画面，感知战场态势变化。在实战化的军事训练环境中，无人机飞行员所需要面临的一个关键问题，即如何在一个高度机动、混杂、变幻的战场环境中，迅速、准确地发现目标并下达攻击指令。一方面，在现实的作战行动中，特别是在无人机担负重要角色的“定点清除”行动中，恐怖分子往往混迹于普通平民之中，难以通过普通的侦察行为辨析二者的差别；另一方面，整个战场环境存在巨大的流动性，时间的推移给目标打击带来更大的不确定性，彼时的军事目标到此时可能转变为非军事目标。上述因素对无人机飞行员的战场感知能力提出了新的挑战。沙瑞尔指出：“机器在特定的军事应用中可以比人类做得更好。例如，当 C－RAM 自主武器系统遂行作战行动的反应速度与人类操作员相比具有压倒性优势，它就可以用于拦截敌方导弹和火箭的饱和攻击。当来袭的导弹具有更多的智能化特征，比如可以发起‘蜂群攻击’，这些有人类监督的自主防御武器将变得更加重要——而人类操作员的参与程度将会下降。”[①] 由此可见，在强调“首发制穴”和追求武器打击速度的战场，受到生理条件限制的人类逐

① 沙瑞尔. 无人军队：自主武器与未来战争［M］. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：366.

渐沦为作战系统中最弱的一环，人与武器相得益彰的结合模式愈发难以实现。

在生理层面，无人机操控员面临的另一个重大挑战是“反应延迟”问题。长期以来，地面控制站显示屏上的影像总比在无人机作战的实时情况滞后几秒，这是因为无人机的信号要经过通信卫星中转。受“反应延迟”效应的影响，无人机操控员对移动目标的打击变得异常困难。2011 年，一名“阿拉伯半岛基地组织”的高级指挥官坦言，如果他们听到美军无人机从头顶飞过，他们就尽可能快地来回跑动，从而干扰无人机飞行员对目标的准确定位。[①] “反应延迟”现象的存在，从一个侧面反映了无人机操控员大脑与无人机系统结合的效率问题。

睡眠是一种在人类认知功能中扮演基础性角色的活动机制。持续的睡眠剥夺会影响作战人员的身体机能，进而对工作效率产生不利影响。对于无人机操控员而言，超长的工作时间同样意味着生理层面的巨大挑战。例如，“捕食者”无人机在战区的架次平均工作时间为 20～22 小时，机组人员虽然是轮班作业，但是每一班持续时间往往也在 10 小时以上。超长时间的工作意味着对睡眠的剥夺，将影响无人机操控员的警惕性、记忆力和知觉辨别力等重要生理机能指标。虽然美军实验室展开了一系列关于睡眠的生理学和分子层面的专项研究，并在无人机操控员的生理学课程着重设置如何调整生物钟、饮食和休息周期等内容，但是面对长达 12 小时的侦察、警戒与作战任务，无人机操控员仍然难以持续保持精神高度集中，化解因过劳而产生的生理压力。

① MAZZETTI M. 美国的无人机战争［EB/OL］.（2012－07－10）［2022－04－07］. https://world.huanqiu.com/article/9CaKrnJw9Re.

另一方面，就心理层面而言，智能化战争模式也给予作战人员在心理层面的新挑战。美国著名智库布鲁金斯学会高级研究员辛格博士在一份研究报告中指出：在阿富汗战场，无人作战部队的作战精神压力远远高于其他作战部队。他们具有更为明显的身心困顿和精疲力竭感。[①] 相比于传统飞行员，无人机飞行员需要具有更过硬的心理素质。传统飞行员由于距离和高度的原因，往往很难真切地看到目标毁伤效果，而无人机飞行员可以直观地看到血肉横飞的画面，这会造成比传统飞行员更为严重的道德压力和责任压力。事实上，自2006年美军开始培养“全球鹰”“捕食者”等无人机的专职飞行员时，就已经关注到操控员面临的心理压力问题，然而对于如何化解这些问题一直未能找到最有效的路径。究其原因，这种特殊的心理压力主要是由于无人作战独特的作战方式。区别于传统战争的作战人员，无人机飞行员需要在虚拟作战环境下操控远在千里之外的作战平台，他们在密布屏幕、按钮和开关的环境下工作，通过远程静音视频遥控操作，感受不到真实战场的声音和气味。这种独特的作战模式导致作战人员在感官上与现实战场的分离，长期的分离与隔绝会产生一种疏离感，进而逐渐引发精神上的分裂。

在心理层面，承担智能化作战任务的作战人员不仅要面对现实战场与虚拟战场分离的心理压力，而且要面对作战与日常生活之间频繁切换带来的额外心理压力。传统战场上，军人战时驻扎在远离家人的军营之中，而在无人作战模式下却截然不同。无人机操控员还面临工作与生活情景切换模式下的心理压力。无人机操控员不同于传统飞行员之处，还体现在他们不用离开家庭就能参加战争的工

① 张煌，傅中力，林聪榕.“生理·心理·物理”三位一体——美军无人作战装备训练模式解析［J］.装备学院学报，2015，26（3）：8.

作方式。在美国内华达州无人机部队任职的马丁中校坦言："在这里，每天上下班就是上下战场，感觉完全不同。晚上6点到早上6点，你可能会先接孩子放学回家，然后路上顺便买点杂货，回家再做顿饭。但是，早上6点到晚上6点这段时间里，你又成了'合法杀手'。"① 长期经历这种通勤方式，令马丁中校等一批无人机飞行员出现精神分裂的症状。更值得注意的是，智能化作战平台的操控者能够在虚拟作战环境中感受到自己现实生活的场景。例如，在执行"定点清除"恐怖分子的作战任务时，无人机飞行员需要长时间监控武装分子，从而能够看到大量恐怖分子与家人一起生活的日常场景，某些家庭生活的温情场景会使飞行员的心理产生微妙的变化，进而可能影响其执行任务的决心。正如美国心理学家奥尔特加所指出的："在某些特定的时刻，无人机飞行员看到的场景可能会让自己回忆起曾经做的一些类似的事情。当一种似曾相识的熟悉感涌上心头，可能导致飞行员在扣动武器系统开关的时候增加心理上的难度。"②

正如上文所述，人类固有的生理、心理局限，导致其已经成为无人机反恐作战系统的"短板"。

2. 机器人的"智能化"挑战

无人作战模式下，高度智能化的机器人既是战争的工具，又可能成为战争的主体。当前，以美国、俄罗斯、以色列为代表的西方军事强国加速推动智能化自主武器系统的军备竞赛。对于智能化的机器人能否真正取代人类成为战争的主体，当前学界存在不同的看

① Sally. 走近美军无人机"飞行员"的生活 [J]. 看世界，2017 (4)：58.

② Plageja. 711hpw/HP research, analysis, and consultation for USA, 711th human performance wing [R]. Air force Research Laboratory, USAF Materials Command, 2010：75 - 146.

法。技术乐观主义的观点认为，具有高度智能化水平的机器人比人类更适合成为未来战争的主体。阿金指出，战争中机器人具有六项人类不可比拟的优势。

与之相对应，也有学者认为战场机器人存在失控的风险，因而对其未来前景持悲观态度。沙瑞尔指出：“这种自主武器在军事上具有重要的价值，同时也会带来潜在的风险。在这些场景下，一旦把任务赋予完全自主武器，如果它失败了，或遭到黑客侵入，或被操控攻击错误的目标，则不能及时将自主武器召回或中止任务。……考虑到这种完全自主武器使用的潜在风险，我们应该对其军事价值和风险进行仔细权衡。”[①] 由于完全自主的机器人可以独立进行战场决策，缺乏人类理性判断和规制，可能导致“意外战争”和危机“意外升级”的风险。

此外，以西班牙巴塞罗那大学的古森兹维格为代表的一些学者认为，智能化机器人难以有效遵循战争的人道主义原则，进而可能增加交战过程中违背人道的行为。[②] 一方面，就遵守区分原则而言，机器人难以准确区分军人和平民，特别是难以判断穿着贫民服饰的恐怖分子，因而存在滥杀无辜的风险。加拿大多伦多大学的卡尔皮教授也指出：机器人很难正确识别对方军人的投降信号，因为感知传递投降信号存在极大的主观性。[③] 另一方面，就遵守相称性原则而言，机器人虽然具有超出常人的数据运算和处理能力，但对于评估军事行为在道德层面的收益是否与其损失相称的问题，机器人很难

① 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：370.

② GUERSENZVAIG A. Autonomous weapon systems: failing the principle of discrimination [J]. IEEE Technology and Society Magazine, 2018, 37 (3): 55-61.

③ KARPPI T, BHLEN M, GRANATA Y. Killer robots as cultural techniques [J]. International Journal of Cultural Studies, 2018, 21 (2): 107-123.

采取客观公正的算法应对复杂多变的道德评估场景。可以这样设想，一旦战争由冰冷的机器人来主宰，战争中机器人所犯下的罪行没有人类为之承担责任，那么战争必然毫无人性可言。

（二）人的机器化增强

相对于人工智能技术驱动作战机器人一日千里的发展，自然人在生理、心理层面的能力素质未能同步提升。为了应对军事智能化时代人与机器的融合问题，进一步提升体系作战能力，当前一个重要的思路就是利用增强技术提升人类的各项能力指标，使之适应和匹配智能化机器发展的需求。

1. 物理增强

物理层面的增强方式是人类趋向机器化增强最显而易见的方式。当前，世界军事强国在物理增强方面的一种主要尝试，即尝试以加装“机器外骨骼”的方式来实现作战人员的体力倍增，进而打造“机甲战士”。早在2012年，为了应对阿富汗战场复杂地形环境给美军机动作战和后勤运输带来的挑战，美国陆军委托国内军工巨头洛克希德·马丁公司测试研发“人体负重外骨骼”系统，用于提高美军士兵在高海拔山地环境的负重能力和运动能力。根据相关研发机构公布的文件显示，“人体负重外骨骼”系统由液压驱动，可以使士兵在任何地形环境中承受200磅（约91千克）的重物，负载的重量通过外骨骼的靴子传递到地面，从而确保重量不会转移到士兵的身体上，不会妨碍士兵的活动。①

① Army Technology. Human Universal Load Carrier (HULC) [EB/OL]. (2020-10-26) [2022-04-07]. https://www.army-technology.com/projects/human-universal-load-carrier-hulc.

2. 生理增强

生理增强的一种重要手段，就是利用脑机接口、生物信息芯片等技术手段来强化作战人员的感知力、记忆力及反应能力，以应对智能化作战中复杂巨系统、海量数据信息及快速多变的特殊环境。自20世纪70年代以来，作为美国国防科技发展的重要引擎，美国国防高级研究计划局持续资助脑机接口领域的科学创新研究和技术开发，在利用脑机接口技术恢复作战人员神经和行为功能及改善作战人员训练效果和作战效能方面取得一些重要进展。[①] 2011年，俄罗斯军方提出了“2045计划”，其中提出加快攻克涉及“脑控”“仿脑”领域的核心关键技术，其目的就是尝试从生理层面增强作战人员能力。与之相对应，美国五角大楼也在2013年提出了“阿凡达”计划，旨在建立人脑远程操控的机器人部队，替代人类进行战场厮杀。[②]

3. 心理增强

无人作战独特的作战环境和作战模式，容易使身处其中的作战人员产生焦虑障碍、精神分裂、自杀倾向，造成心理疾病。为了应对无人作战环境对于军人心理层面的新挑战，美国空军航天医学中心成立了专门的项目研究小组，对无人机飞行员及技术支持人员所承受的心理压力展开专项调研。他们认为，要确保无人机飞行员在高强度的远程虚拟作战中规避现实生活环境的潜在影响，就必须使其具备类似狙击手的心理素质，即能够在经历单调冗长的等待后立

① MIRANDA R A, CASEBEER W D, HEIN A M, et al. DARPA - funded efforts in the development of novel brain—computer interface technologies [J]. Journal of Neuroscience Methods, 2015, 244 (15): 52 - 67.

② 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：8.

即做出高水平的快速反应和精准决策。为此，美军一方面运用明尼苏达多项人格测验、冯德里克人员测试和迈尔斯－布里格斯性格测试等专业方法进行心理评估，找到无人机飞行员的心理短板；[①] 另一方面，为了确保无人机飞行员心理素质的稳定性，2008 年以来，美军心理战部队不断探索心理增强的技术手段创新，尝试将心理医学中的“正念减压疗法”引入心理增强领域。通过有针对性的心理训练，强化内在心理弹性，以此帮助无人机飞行员调节和管理应激情况对身心冲击造成的诸多不良影响，同时提高无人机飞行员的记忆力和专注力，扩展对于作战数据的记忆容量，最大程度降低无人作战带来的认知损耗，进一步提升认知效率。

（三）机器的人化设计

进入21 世纪，智能化机器人技术的进步日新月异，各类无人作战平台的军事应用呈现井喷式发展。以美军为例，从在反恐战场上独领风骚的“捕食者”无人机，到在阿富汗战场上执行清除简易爆炸装置任务的履带式魔爪机器人，再到波士顿动力学工程公司研制的“大狗”负重运输机器人，机器人已经渗透到美军作战和军事行动的方方面面。在叙利亚战场上，俄军已经组建了整建制的机器人部队，可以承担情报侦察、武装突袭、排雷爆破等多样化作战任务。

随着机器人智能化程度的不断提升，机器人作为未来战争重要的战争主体的发展趋势日益明显，而关于机器人可能摆脱人类控制、沦为杀戮工具和使得战争彻底失去人性的忧虑也日渐凸显。人类与机器人如何能够在未来的智能化战争中做到团结协作？如何规避机

① HOAGLAND B T. Manning the next unmanned air force: developing RPA pilots of the future [R]. Brookings, 2013: 1.

器人战争走向失控的风险？这一系列问题的解决既需要人类在物理、生理、心理层面能力素质的提升，也需要在机器人的设计研发过程中谨守人类控制底线，彰显人道主义的精神。因此，人机深度融合必须实现机器人武器装备的技术设计与伦理设计的同步并行，具体伦理设计的原则体现为以下三个方面。

1. 坚守人类控制机器的设计底线

虽然当前学界仍有少数学者对机器人作为独立的战争主体持乐观态度，但绝大部分专家学者对于发展完全自主的机器人持谨慎态度。在2018年的国际人工智能联合会议上，2000多名人工智能领域的专家共同签署《致命性自主武器宣言》，倡议禁止研发具有致命能力的自主机器人系统。[①] 在这些人工智能专家看来，必须将未来战场机器人的设计置于人类可以控制的范围之内。那么，如何在具体设计过程中坚持人类控制的底线？笔者认为，人类可控的设计原则主要应当体现为以下三个方面：一是主动的控制，即在整个作战系统的设计中，人机关系应当被设计为人类驾驭机器，而非机器驾驭人类，人类不能被机器反向操控，人类应当积极主动地应对未来战场机器人的潜在威胁，而不能被动地将人类的控制当作机器人作战失控后的补救措施；二是全过程的控制，即必须保证将人类主动控制的权力贯穿于战场机器人设计、研制、开发、生产乃至应用的整个流程；三是可理解的控制，即必须确保机器人的设计过程规避人工智能的“技术黑箱”，设计者不仅要充分理解设计过程中算法规则与运算过程，还必须深入理解每项算法运算在道德层面可能导致的潜在后果。

① Future of Life Institute. Lethal autonomous weapons pledge [EB/OL]. (2018-06-06) [2022-04-07]. https://futureoflife.org/lethal-autonomous-weapons-pledge/.

2. 实现自上而下的道德预置

人类对于技术的设计过程同样也是价值观念植入的过程。因此，可以在机器人的设计过程中预先植入战争伦理价值观念，使得机器人在作战过程中能够按照预先设置的道德观念来执行作战任务。那么，机器人应当如何遵循道德规范和伦理原则呢？目前，人工智能领域学者考虑以自上而下的方式实现道德预置。阿金认为，我们可以通过给自主武器编程，使之永不违反战争法。它们根本无法违反战争法。它们不会寻求报复，不会感到生气或恐惧，丝毫不会受到任何情绪的影响。它们能在必要时进行杀戮，又能在行为非法时立刻终止。[①] 具体而言，就是利用计算机算法将伦理原则转化为机器人能够理解的作战规则。例如，依据正义战争伦理观念，交战各方要遵循区分原则，即严格区分军事人员和非军事人员，确保平民享有军事打击的豁免权。为了使机器人遵循区分原则，可以采用自上而下设置一组关于区分的算法，用于帮助机器人区分军人和平民。虽然也有一些学者对自上而下的道德预置方法提出批评，认为当前的算法只能将军服、武器等标识物设置为区分目标，对于身穿平民服饰的恐怖分子却束手无策。但是，作为机器人趋向“人化”的一种努力，随着人工智能水平的不断提升，这种道德预置的方式可能成为解决人机融合风险挑战的一条重要的进路。

3. 实现自下而上的道德生成

在将人类的道德观念嵌入机器人的过程中，必须坚持自下而上的道德生成与自上而下的道德预置相互结合。考虑到当前人工智能技术的发展水平，依靠算法实现自上而下的道德预置仍然存在较大

① 沙瑞尔. 无人军队：自主武器与未来战争［M］. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：320.

的技术“盲区”和不确定性。为此，阿金提出，充分利用当前人工智能领域在智能学习、机器学习和深度学习领域取得的一系列前沿进展，将其应用于作战目标的区分与识别。[①] 具体而言，在自上而下的道德预置的基础上，对机器人在作战过程中的道德表现进行长时间的样本观察和数据积累，利用“改进—继续—改进”的机器学习模式，不断修正算法，并改进机器人在作战过程中的道德表现，依靠机器人的自我纠正来实现执行伦理原则能力的持续提升。这种自下而上的道德生成方式，可以作为自上而下道德预置模式的重要补充，共同推动机器人的自我完善与自我发展，成为规避机器人失控风险和战争“非人化”道德挑战的重要进路。

回顾历史，军事技术进步和作战方式的演变对于人机融合的挑战始终是战争的重要主题。当前，智能化战争模式下的人机融合问题，其本质是由于人类与武器系统的关系发生了革命性变化，呈现从原来的“人在系统之中”，到现在的“人在系统之上”，再到未来的“人在系统之外”的发展趋势。在这一背景下，考虑到机器的智能化、自主化发展趋势不可逆转，人类采取何种方式方法强化自身的生理、心理素质，人类以何种方式方法与机器深度结合，就决定了未来战争中战斗力的生成和提升的空间。

在传统战争模式下，军人必须比平民承担更大的战争风险，这正是军人职业存在的价值所在。面对全球军事智能化如火如荼的浪潮，军人应当具有强烈的危机感和紧迫感。在未来机器人相互厮杀的战场环境下，军人处在比平民更为安全的环境下遥控指挥战场，如何继续履行战争主体责任？如何继续展现出自身的价值？如何继

① ARKIN R C. The case for ethical autonomy in unmanned systems [J]. Journal of Military Ethics, 2010, 9 (4): 333.

行动中，针对苏莱曼尼的刺杀行动具有独特性和里程碑意义。刺杀苏莱曼尼是美国首次以反恐为名针对主权国家的重要军事领导人实施的无人机“定点清除”行动。相比于此前针对“基地”组织、“伊斯兰国”等恐怖组织领导人和恐怖分子展开的“定点清除”行动，此次刺杀行动的目标身份截然不同。正如美国《时代》周刊所指出的：“苏莱曼尼是受国际广泛关注的中东国家领导人和军事战略家，也是伊朗区域外交政策的公开面孔和仲裁者，与黎巴嫩、伊拉克、叙利亚、阿富汗和也门等中东各地的非国家行为体建立了牢固的关系。……自美国于2018年5月退出伊朗核协议以来，德黑兰和华盛顿经历了不断升级的紧张局势和相互挑衅，杀死苏莱曼尼是华盛顿迄今为止最强烈的信号，表明其行动规则已经改变。”[①] 特朗普政府通过击毙苏莱曼尼的军事行动，将其反恐矛头直接指向“国家恐怖主义”及其“代言人”，这一反恐行动规则的转变，使得此次刺杀事件在正义性层面面临更多和更为尖锐的争议。

同时，美军刺杀苏莱曼尼事件也是在总统直接指挥下、由“察打一体”无人机完成的反恐刺杀事件，是美国在全球范围内发动“反恐战争”中具有国际影响力的案例之一。对这一事件展开深度的案例分析，有助于将美国“反恐战争”正义性探讨从宏观层面进一步向微观层面推进。回顾历史，自“9·11”事件和美国发起阿富汗战争以来，“察打一体”的武装无人机日益成为美国反恐作战的重要力量，特别是在对恐怖组织头目的“定点清除”行动中扮演着至关重要的角色，由此引发一系列伦理和法理问题。国内外学者围绕美军无人机反恐作战正义性的探讨也随之展开，相关研究成果日益丰

① VAKIL S. Qasem Soleimani's assassination opens a Pandora's Box in the Middle East [J]. Time, 2020: 1.

第六章

案例分析：以美军刺杀苏莱曼尼为例

2020年1月3日，美国总统特朗普向外界公布：1月2日晚，在他的直接领导下，美军使用无人机对离开伊拉克巴格达国际机场的两辆车辆实施“定点清除”式打击，击毙了伊朗“圣城旅”指挥官苏莱曼尼。[①] 美军刺杀苏莱曼尼的消息一经公布，便引发国际舆论的高度关注。国际主流媒体、智库和国际政治、国际法领域的专家学者积极发声，从政治、经济、军事、宗教等不同角度，分析刺杀事件发生的原因并预测其可能造成的潜在影响。

在围绕刺杀事件展开的诸多讨论中，正当性与合法性问题成为其中的一个焦点议题。在美国政府发起的一系列无人机“定点清除”

① National Security & Defense. Remarks by President Trump on the killing of Qasem Soleimani [EB/OL]. (2020-01-04) [2022-04-07]. https://user.guancha.cn/main/content?id=221304&s=fwzxfbbt.

一方面，特朗普政府声称刺杀苏莱曼尼是针对其恐怖主义行径的自卫行为。总体而言，历届美国政府对其“定点清除”反恐行动的正义性辩护，都是以遵循自卫原则这一正当理由来展开的。美国政府坚持认为，恐怖组织不但策划制造了“9·11”事件，而且实施了一系列针对美国的恐怖主义袭击，严重威胁和损害了美国国家安全。美国发动阿富汗战争打击塔利班和“基地”组织，属于行使自卫权利的正当行为。在恐怖主义势力向全球蔓延并对美国国家利益构成持续威胁之时，这一理由依旧有效。美国政府和军方在全球范围内使用无人机“定点清除”恐怖分子头目的军事行动，属于自卫权利的自然延伸，从而给予特朗普政府刺杀苏莱曼尼的军事行动以正当理由。事实上，早在2007年，美国政府就宣布苏莱曼尼领导的“圣城旅”部队为恐怖主义的支持者；到2017年4月，美国又进一步出台国内法规，将“圣城旅”部队直接认定为外国恐怖主义组织。[①] 刺杀行动发生之后，美国总统特朗普发表讲话，将苏莱曼尼认定为“世界上第一大恐怖分子”，“苏莱曼尼在过去20年中一直在进行恐怖主义行为，破坏了中东的稳定”，“苏莱曼尼助长了从新德里直至伦敦的恐怖主义阴谋，使无辜人民成为他病态激情的牺牲品”。[②] 在美国国防部发表的声明中，也明确将自卫作为刺杀行动的正当理由，强调“美国军方采取了果断的防卫行动，通过杀死美国指定的外国恐怖主义组织伊斯兰革命卫队‘圣城旅’负责人苏莱曼

① GALBRAITH J. U. S. Drone strike in Iraq kills iranian military leader Qasem Soleiman [J]. American Journal of International Law, 2020, 114 (2): 313-323.

② National Security & Defense. Remarks by President Trump on the killing of Qasem Soleimani [EB/OL]. (2020-01-04) [2022-04-07]. https://user.guancha.cn/main/content?id=221304&s=fwzxfbbt.

富。总体而言，目前各国学者围绕这一议题的探讨，主要是在宏观层面将美军无人机反恐作为一个整体性事件，系统考察无人机反恐对传统正义战争理论体系的冲击与挑战。就宏观层面而言，相关分析探讨的系统性、全面性是毋庸置疑的。但就微观层面而言，考虑到不同总统任期内美国反恐政策存在一定的差异性，无人机反恐在不同时间、地域、交战环境下的作战运用呈现出多样性特征，由此引发伦理、法理层面的探讨也不尽相同。当前学界缺乏对美国无人机反恐作战正义性的深度案例剖析，因此造成的研究空白以及研究成果的局限性也是显而易见的。

本章从正义战争理论的分析框架出发，系统梳理特朗普政府对于刺杀苏莱曼尼的事件的正义性陈述，细致考察国际社会对于此次刺杀行动的一系列争议，依据开战正义的正当理由原则和交战正义的区分原则、相称性原则等伦理评判标准，对特朗普政府刺杀苏莱曼尼的军事行动从正义性层面进行评判，深入揭示刺杀行动在遵循开战正义、交战正义原则方面存在的问题，进一步揭示导致刺杀事件正义性争议的深层次原因。

一、美国定义刺杀事件的“正当理由”

在美军刺杀苏莱曼尼事件发生以后，美国政府、军方纷纷发声，试图为这场刺杀行动进行辩护和粉饰，强调美国具有无可争议的正当性与合法性，力图占据法律和道义上的制高点，赢得国际舆论的主动权。美国总统特朗普的讲话、国防部发布的声明以及美国常驻联合国代表致安全理事会主席的信，主要从遵循自卫原则和行使预先防卫权两个方面，对使用无人机刺杀苏莱曼尼事件的“正当理由”进行陈述。

一方面，特朗普政府声称刺杀苏莱曼尼是针对其恐怖主义行径的自卫行为。总体而言，历届美国政府对其“定点清除”反恐行动的正义性辩护，都是以遵循自卫原则这一正当理由来展开的。美国政府坚持认为，恐怖组织不但策划制造了“9·11”事件，而且实施了一系列针对美国的恐怖主义袭击，严重威胁和损害了美国国家安全。美国发动阿富汗战争打击塔利班和“基地”组织，属于行使自卫权利的正当行为。在恐怖主义势力向全球蔓延并对美国国家利益构成持续威胁之时，这一理由依旧有效。美国政府和军方在全球范围内使用无人机“定点清除”恐怖分子头目的军事行动，属于自卫权利的自然延伸，从而给予特朗普政府刺杀苏莱曼尼的军事行动以正当理由。事实上，早在2007年，美国政府就宣布苏莱曼尼领导的“圣城旅”部队为恐怖主义的支持者；到2017年4月，美国又进一步出台国内法规，将“圣城旅”部队直接认定为外国恐怖主义组织。[①] 刺杀行动发生之后，美国总统特朗普发表讲话，将苏莱曼尼认定为“世界上第一大恐怖分子”，“苏莱曼尼在过去20年中一直在进行恐怖主义行为，破坏了中东的稳定”，“苏莱曼尼助长了从新德里直至伦敦的恐怖主义阴谋，使无辜人民成为他病态激情的牺牲品”。[②] 在美国国防部发表的声明中，也明确将自卫作为刺杀行动的正当理由，强调“美国军方采取了果断的防卫行动，通过杀死美国指定的外国恐怖主义组织伊斯兰革命卫队‘圣城旅’负责人苏莱曼

① GALBRAITH J. U. S. Drone strike in Iraq kills iranian military leader Qasem Soleiman [J]. American Journal of International Law, 2020, 114 (2): 313-323.

② National Security & Defense. Remarks by President Trump on the killing of Qasem Soleimani [EB/OL]. (2020-01-04) [2022-04-07]. https://user.guancha.cn/main/content?id=221304&s=fwzxfbbt.

尼，来保护美国在国外的人员”。[①] 此后，美国常驻联合国代表克拉夫特致信联合国安全理事会主席，再一次重申了这一理由：“美国根据《联合国宪章》第51条，为行使其固有的自卫权采取了某些行动，这些行动是针对最近几个月伊朗伊斯兰共和国和伊朗支持的民兵对美国军队和中东地区利益的一系列武装袭击的反应。”[②] 在上述陈述中，美国对实施刺杀行动给出的理由是高度一致的，即针对苏莱曼尼领导的恐怖主义袭击进行自卫还击。

另一方面，特朗普政府又声称刺杀苏莱曼尼的行动是针对“迫在眉睫的威胁”，迫切需要美国行使预先防卫权。在正义战争理论中，履行自卫权利是实现开战正义的唯一正当理由，当代正义战争理论家进一步拓展了对正当理由原则的诠释，主张国家在面临迫在眉睫的军事威胁时，有权利以先发制人的方式进行预先防卫。特朗普政府强调苏莱曼尼正在策划新一轮针对美国的恐怖袭击，从而为刺杀行动的正义性辩护寻找到更为丰富的正当理由。特朗普指出：“苏莱曼尼正在筹划对美国外交官和军事人员进行迫在眉睫的险恶袭击，但我们将他锁定，并击毙了他。”[③] 美国国防部也在官方声明中给出了与之相类似的理由“苏莱曼尼将军正在积极制订计划，袭击伊拉克和整个中东地区的美国外交官和服务人员”，并强调“这次刺

① U. S. Department of Defense. Statement by the Department of Defense [EB/OL]. (2020 - 01 - 02) [2022 - 04 - 07]. https://www.defense.gov/News/Releases/Release/Article/2049534/statement-by-the-department-of-defense/.

② Letter dated 8 January 2020 from the Permanent Representative of the United States of America to the United Nations addressed to the President of the Security Council [Z]. United Nations Digital Library, 2020.

③ National Security & Defense. Remarks by President Trump on the Killing of Qasem Soleimani [EB/OL]. (2020 - 01 - 04) [2022 - 04 - 07]. https://user.guancha.cn/main/content?id=221304&s=fwzxfbbt.

杀行动旨在阻止未来伊朗的进攻计划”。[①] 美国国防部长埃斯珀进一步指出：“如果我们听到攻击的消息，我们将采取先发制人的行动，以保护美军和美国公民生命。”[②] 在上述陈述中，特朗普政府和军方刻意使用时间层面的“迫在眉睫”、空间层面的覆盖“伊拉克和整个中东地区”，以及程度层面的“险恶的袭击”等词语，渲染出潜在军事威胁的紧迫性和严峻性，从而显示出行使预先防卫权利的正当性。按照特朗普政府的逻辑，美国面临恐怖袭击的威胁越大，实施预先防卫行动的迫切性也就越大，因此，美国必须以先发制人的方式，清除潜在的战争威胁，以避免危机进一步升级。

值得注意的是，特朗普政府不仅宣称美国具有刺杀苏莱曼尼的正当理由，还力图证明整个军事行动的正义性，为刺杀行动全过程进行道德辩护。依据特朗普本人以及美国政府、军方对刺杀事件的官方陈述，刺杀苏莱曼尼的意图并不仅仅是为了美国的利益，而是为了“防止进一步危害国际和平与安全的升级事件”，[③]，以及“寻求与世界所有国家的和平、和谐与友谊”。[④] 刺杀苏莱曼尼的过程是用无人机成功完成的“一次完美无瑕的精确打击”，[⑤] 严格遵循区分

① U. S. Department of Defense. Statement by the Department of Defense [EB/OL]. (2020 - 01 - 02) [2022 - 04 - 07]. https://www.defense.gov/News/Releases/Release/Article/2049534/statement-by-the-department-of-defense/.

② CROWLEY M, HASSAN F, SCHMITT E. U. S. Strike in Iraq kills Qassim Suleiman, commander of Iranian force [N]. New York Times, 2020 - 01 - 02.

③ Letter dated 8 January 2020 from the Permanent Representative of the United States of America to the United Nations addressed to the President of the Security Council [Z]. United Nations Digital Library, 2020.

④ National Security & Defense. Remarks by President Trump on the killing of Qasem Soleimani [EB/OL]. (2020 - 01 - 04) [2022 - 04 - 07]. https://user.guancha.cn/main/content?id=221304&s=fwzxfbbt.

⑤ National Security & Defense. Remarks by President Trump on the killing of Qasem Soleimani [EB/OL]. (2020 - 01 - 04) [2022 - 04 - 07]. https://user.guancha.cn/main/content?id=221304&s=fwzxfbbt.

原则和相称性原则。对于刺杀行动带来的影响，特朗普宣称“我们昨晚采取行动制止了战争，我们没有采取行动发动战争”，“我们为苏莱曼尼的恐怖统治已经结束感到安慰”，并坚信“没有这些怪物，世界是一个更安全的地方”。[①] 由此可知，当代美国的政治家和战略家们对于正义战争理论、伦理原则和规范的认知与把握能力，丝毫不逊色于现代战争伦理和国际法领域的专家学者，他们能够务实、娴熟地运用现有伦理和法律理论框架，论证美国对外军事行动的合法性与正当性。

二、刺杀事件是否遵循开战正义的伦理原则？

在正义战争理论体系中，开战正义主要探讨国家诉诸战争及其他低烈度军事行动的权利问题，用于衡量政治领导人发动战争的决策是否具有正当性和合法性，其中，正当理由原则是评判军事行动是否符合开战正义标准的核心原则。因此，评判美国无人机刺杀苏莱曼尼是否符合开战正义要求，主要看其是否具有正当理由。正如上文所述，特朗普和美国政府就美军杀害苏莱曼尼给出了两个方面的理由：一方面是基于自卫的理由，另一方面是基于预先防卫的理由。然而，依据正义战争理论关于正当理由原则的界定来分析，特朗普政府给出的两方面理由都存在较大的争议。

一方面，对于特朗普政府提出行使自卫权的开战理由，正义战争理论明确规定，国家基于自我防卫而进行武力反抗虽然是正当的行为，但行使自卫权利的前提条件是遭遇他国的武装侵略，国家主

① National Security & Defense. Remarks by President Trump on the killing of Qasem Soleimani [EB/OL]. (2020 - 01 - 04) [2022 - 04 - 07]. https://user.guancha.cn/main/content?id=221304&s=fwzxfbbt.

而在于面临威胁的充分程度，并提出了评估威胁充分程度的三个条件。[①] 依据沃尔泽提出的标准，我们可以对美国以预先防卫为由刺杀苏莱曼尼的正义性进行评估。

第一个条件，公开表明实施伤害的意图。沃尔泽认为，对于敌方实施伤害的意图的判断，不能够依据潜在的情况或是己方的主观臆测，而应当依据对方的公开表述或已摆出的进攻姿态。虽然特朗普在声明中公开指责："多年来，在苏莱曼尼的领导下，伊斯兰革命卫队及其残酷的'圣城旅'已经伤害并谋杀了数百名美国平民和军人。"[②] 但是，一方面，特朗普没有为上述指控提供真凭实据，另一方面，苏莱曼尼本人没有公开发表对美国实施军事打击的言论。与之相反，苏莱曼尼在写给美国驻伊拉克大使的信件中坦言："我在霍梅尼的坟墓上发誓，我没有授权对美国发射'子弹'。"英国记者梅拉就此分析指出，苏莱曼尼可能就美国的中东政策提出抗议，但是他的克制能力是真实的，进攻美国是有风险的。[③] 作为伊朗军方的头面人物，苏莱曼尼公开发表攻击美国的言论并非明智之举，也是与伊朗国家利益背道而驰的。

第二个条件，积极的准备使得这种伤害的意图具有转变为现实的危险。在沃尔泽看来，这一评估标准主要涵盖两方面的因素：一方面是"积极的准备"，另一方面是"现实的威胁"。对于衡量威胁的充分程度而言，这两方面的因素缺一不可。"积极的准备"并不等同于造成了"现实的威胁"，两者之间没有必然的因果联系。一国在

① 沃尔泽. 正义与非正义战争：通过历史实例的道德论证［M］. 任辉献，译. 南京：江苏人民出版社，2008：92.

② National Security & Defense. Remarks by President Trump on the killing of Qasem Soleimani［EB/OL］.（2020－01－04）［2022－04－07］. https://user.guancha.cn/main/content?id=221304&s=fwzxfbbt.

③ MERAT A. The commander［J］. Prospect, 2019（282）：34－38.

其中涉及苏莱曼尼对美国所犯的罪行体现为："由苏莱曼尼指挥的'圣城旅'部队，支持了包括'卡塔布真主党'在内的伊拉克民兵武装，已针对美国驻伊拉克部队所在地进行了一系列间接的火力袭击。2019 年 12 月 27 日，一次此类袭击导致美国政府承包商死亡，四名美国军人受伤。"① 克拉夫特提出的罪责虽然具体，但是他未能就苏莱曼尼与伊拉克国内什叶派民兵武装之间的关联性给出直接证据，未能证明苏莱曼尼及"圣城旅"部队直接参与了对美武装袭击。同时，针对美方的指责，伊朗常驻联合国代表矢口否认伊朗军方对伊拉克民兵组织的支持，并声称"绝对拒绝美国将有争议的民兵组织袭击行动归因于伊朗"。② 因此，特朗普政府将自卫行动的目标瞄准伊朗军方和苏莱曼尼，难以得到国际社会的普遍认可。

另一方面，对于特朗普政府提出行使预先防卫权的理由，一个较早的评判标准来自美国国务卿维博斯特提出的观点，即只有存在"刻不容缓、压倒一切的、没有选择方式的余地、没有时间深思熟虑的自卫需要"，才能够证明行使预先防卫权利的正当性。③ 然而，在沃尔泽等当代战争伦理学家看来，这种观点所界定的"预先防卫"，就像是一种本能的反射性动作，其行动的正当性几乎无须证明。沃尔泽指出，由于威胁发生难以从时间层面进行准确预测，因此，确定预先防卫是否具有正当性的关键，不在于面临威胁的紧迫程度，

① Letter dated 8 January 2020 from the Permanent Representative of the United States of America to the United Nations addressed to the President of the Security Council [Z]. United Nations Digital Library, 2020.

② Letter dated 29 January 2020 from the Permanent Representative of the Islamic Republic of Iran to the United Nations addressed to the Secretary – General and the President of the Security Council [Z]. United Nations Digital Library, 2020.

③ 沃尔泽. 正义与非正义战争：通过历史实例的道德论证 [M]. 任辉献，译. 南京：江苏人民出版社，2008：84 – 85.

军事合作，共同打击逊尼派叛乱分子的部队。[①] 三是就准备的程度而言，尚没有逾越从威慑到威胁的底线。英国《展望》杂志撰文指出："苏莱曼尼指挥的'圣城旅'部队是一支精锐的准军事部队，直接向伊朗最高领导人汇报。他们搜集情报，建立政治联盟，并为反对德黑兰的外国军事力量创造敌对环境。"[②] 由此可见，苏莱曼尼及其军事力量主要通过搜集情报、援助装备、舆论造势以及建立政治联盟等方式，对其盟友提供隐性支持，这仅能视为一种有约束的对立，一种在限度之内表示抗议的方式，尚未突破到实质性军事威胁的程度。

第三个条件，行动迟缓或采取武力以外的其他手段，将会极大地加剧面临威胁的整体形势。依据正义战争理论，一国针对外部威胁的应对手段是多样化的，只有在谈判、制裁、遏制等手段被穷尽，而外部军事威胁仍在不断加剧的情况下，以军事手段行使预先防卫权才具有正当性。对于美国刺杀苏莱曼尼的军事行动而言，一方面，以非军事手段缓解美伊之间紧张局势的可能性始终存在。自 2018 年 5 月美国宣布退出伊核协议以来，美伊关系持续恶化，然而，即便是在 2019 年 11 月 5 日，伊朗宣布重启位于福尔多铀浓缩厂的离心机之时，总统鲁哈尼仍然强调，尽管伊朗分阶段中止了部分承诺，但仍然致力于与一些国家进行幕后谈判，以寻求维护 2015 年协议的解决方案。[③] 由此可见，在刺杀行动发生之前，美伊之间双边外交磋商谈判的大门并未关闭，英国、法国、俄罗斯和德国等伊核协议缔约国仍在为重启伊核协议积极奔走，联合国原子能机构核查人员重返伊朗的途径未被断绝。另一方面，国际社会和美国主流媒体都在

① MERAT A. The commander [J]. Prospect, 2019 (282): 34 - 38.

② MERAT A. The commander [J]. Prospect, 2019 (282): 34 - 38.

③ AFP Published. Iran steps up uranium enrichment [N]. Global Times, 2019 - 11 - 05.

国防和军队建设领域的积极投入与持续准备，只是具备了某种程度上的威慑能力，并不能够直接判定为具备进攻他国的现实威胁，对威胁的判定必须提供造成实质性伤害的凭据。在美国政府和西方主流媒体眼中，长期以来，苏莱曼尼及其领导的“圣城旅”一直在进行持续而积极的军事准备。《纽约时报》撰文指出：“苏莱曼尼将军是过去二十年来伊朗情报和军事力量几乎所有重要行动的设计师。”① 美国《时代》周刊专栏作家梅拉也撰文指出：“苏莱曼尼与黎巴嫩、伊拉克、叙利亚、阿富汗和也门等中东各地的反美非国家行为体建立了牢固的关系。”② 然而，依据沃尔泽所提出的评判标准，即便苏莱曼尼领导伊朗军事力量正在为危害美国安全利益进行“积极的准备”，但是尚未真正转化为对美国的“现实的威胁”。一是就准备的意图而言，苏莱曼尼进行军事准备的根本目的和出发点是捍卫伊朗的国家安全利益。正如梅拉所指出的，“伊朗军事准备是为了寻求维护领土完整并追求自己的经济和国家安全利益”，而并非围绕某一个具体的战略对手进行扩军备战。二是就准备的方向而言，苏莱曼尼及其军事力量虽然在伊朗周边地区构建了错综复杂的安全网络和联系，但并没有将矛头直接指向美国。梅拉指出，苏莱曼尼对反美主义的热情不及他为伊朗谋求区域利益方面的决心。基于此，在伊拉克，直接目标首先是萨达姆，之后是以“伊斯兰国”为代表的恐怖组织；在叙利亚，苏莱曼尼的目标是各种反政府组织和军事武装；在黎巴嫩，以色列人对苏莱曼尼支持的真主党盟友发动攻击。出于对伊朗国家安全利益的考量，苏莱曼尼甚至一度与美国人展开

① CROWLEY M, HASSAN F, SCHMITT E. U. S. strike in Iraq kills Qassim Suleiman, commander of Iranian Force [N]. New York Times, 2020-01-02.

② VAKIL S. Qasem Soleimani's assassination opens a Pandora's Box in the Middle East [J]. Time, 2020: 1.

并不意味着其可以违反交战正义的伦理规约。在沃尔泽等正义战争理论家看来，开战正义的正当理由缺失，反而对遵循交战正义提出了更高的道义要求。

刺杀行动发生以后，特朗普政府和美国军方不仅宣称美国具有实施刺杀行动的“正当理由”，而且强调美军在刺杀苏莱曼尼过程中充分地遵循了交战正义的伦理原则。特朗普在就刺杀事件发表讲话的开篇即指出：“昨晚，在我的指导下，美国军方成功进行了一次完美无瑕的精确打击，杀死了世界上第一大恐怖分子卡西姆·苏莱曼尼”。[①] 显而易见，特朗普试图用“一次完美无瑕的精确打击”来进一步为刺杀行动正名。然而，依据交战正义的区分原则和相称性原则，在刺杀苏莱曼尼过程中，美军是否真正遵循了交战正义的伦理要求，做到了道义上“完美无瑕”，同样也存在较大的争议。

首先是遵循区分原则方面存在争议。区分原则强调非战斗人员豁免权，规定武力使用过程中应当在平民与战斗人员之间、民用设施与军用设施之间加以区别，不应对民众生活基础以及生存环境造成不必要的破坏。依据区分原则，施加武力应仅以战斗人员和军事设施为目标，而绝不能将平民和民用设施作为蓄意攻击的对象。然而，这一原则并未完全禁止因攻击“合法目标”而造成的“附带伤害”。因此，在遵循区分原则下产生“附带伤害”的道义限度问题，也成为近年来围绕无人机反恐的一个伦理争议焦点。

美国刺杀苏莱曼尼的军事行动是否符合区分原则，存在多方面的争议。其一，苏莱曼尼是否属于合法目标存在争议。特朗普政府将伊朗“圣城旅”部队定义为“恐怖组织”，并将其领导者苏莱曼

① National Security & Defense. Remarks by President Trump on the killing of Qasem Soleimani [EB/OL]. (2020-01-04) [2022-04-07]. https://user.guancha.cn/main/content?id=221304&s=fwzxfbbt.

不同程度上认为，美伊之间的安全环境恶化程度，远未达到行使军事手段所要求的极限。美国《时代》周刊指出："自美国于2018年5月退出伊核协议以来，德黑兰和华盛顿经历了不断升级的紧张局势和相互挑衅，但他们通过尊重彼此的红线避免了直接冲突。"① 《纽约时报》也撰文指出："美国和伊朗长期以来在包括伊拉克、也门和叙利亚在内的整个中东战场上卷入了一场'影子战争'。这些策略通常使用代理人进行战斗，从而为华盛顿和德黑兰之间的直接对抗提供了缓冲。"② 以上两个方面都证明，对于刺杀事件发生前的美伊关系而言，以军事手段进行预先防卫远不是最后的选择，经济、外交等非军事手段的作用均未能尝试到最后时刻或发挥到最大限度。

综上所述，在正义战争理论框架下，依据正当理由原则的界定，审视特朗普政府就刺杀行动的正当性所提出的两方面理由，可以得出结论，特朗普政府无论以行使自卫权或是以行使预先防卫权之名实施刺杀行动，都难以满足正当理由原则的要求，故而在开战正义方面是站不住脚的。特朗普政府高调宣扬刺杀行动的正义性和合法性，只是以道义的外衣掩饰强权的事实，本质上是与正义战争的精神背道而驰的。

三、刺杀事件是否遵循交战正义的伦理原则？

在正义战争的理论体系中，开战正义与交战正义在逻辑层面是相互独立的，彼此不构成因果关系。美国发动不正义的刺杀行动，

① VAKIL S. Qasem Soleimani's assassination opens a Pandora's Box in the Middle East [J]. Time, 2020: 1.

② CROWLEY M, HASSAN F, SCHMITT E. U. S. strike in Iraq kills Qassim Suleiman, commander of Iranian Force [N]. New York Times, 2020-01-02.

并不意味着其可以违反交战正义的伦理规约。在沃尔泽等正义战争理论家看来，开战正义的正当理由缺失，反而对遵循交战正义提出了更高的道义要求。

刺杀行动发生以后，特朗普政府和美国军方不仅宣称美国具有实施刺杀行动的“正当理由”，而且强调美军在刺杀苏莱曼尼过程中充分地遵循了交战正义的伦理原则。特朗普在就刺杀事件发表讲话的开篇即指出：“昨晚，在我的指导下，美国军方成功进行了一次完美无瑕的精确打击，杀死了世界上第一大恐怖分子卡西姆·苏莱曼尼”。[①] 显而易见，特朗普试图用“一次完美无瑕的精确打击”来进一步为刺杀行动正名。然而，依据交战正义的区分原则和相称性原则，在刺杀苏莱曼尼过程中，美军是否真正遵循了交战正义的伦理要求，做到了道义上“完美无瑕”，同样也存在较大的争议。

首先是遵循区分原则方面存在争议。区分原则强调非战斗人员豁免权，规定武力使用过程中应当在平民与战斗人员之间、民用设施与军用设施之间加以区别，不应对民众生活基础以及生存环境造成不必要的破坏。依据区分原则，施加武力应仅以战斗人员和军事设施为目标，而绝不能将平民和民用设施作为蓄意攻击的对象。然而，这一原则并未完全禁止因攻击“合法目标”而造成的“附带伤害”。因此，在遵循区分原则下产生“附带伤害”的道义限度问题，也成为近年来围绕无人机反恐的一个伦理争议焦点。

美国刺杀苏莱曼尼的军事行动是否符合区分原则，存在多方面的争议。其一，苏莱曼尼是否属于合法目标存在争议。特朗普政府将伊朗“圣城旅”部队定义为“恐怖组织”，并将其领导者苏莱曼

① National Security & Defense. Remarks by President Trump on the killing of Qasem Soleimani [EB/OL]. (2020-01-04) [2022-04-07]. https://user.guancha.cn/main/content?id=221304&s=fwzxfbbt.

尼指认为“资助国外流血事件的恐怖主义军阀”。对此，伊朗常驻联合国代表并不认同，他指出，“一个国家将另一国武装部队的正式分支机构指定为所谓的外国恐怖主义组织，这违反了公认的国际法原则和《联合国宪章》，包括国家主权平等原则”，并指责美国的做法是“国家恐怖主义”行径。[①] 其二，苏莱曼尼的同行人员是否属于合法目标也存在争议。《纽约时报》报道，美军“收割者”无人机发射多枚“地狱火”导弹攻击苏莱曼尼所在车队，造成同行人员死亡，其中两人身份尚未识别。[②] 即便依据美国的指控，苏莱曼尼在伊拉克境内支持和策划恐怖主义行动，严重威胁中东地区安全，对美国承包商遇害和大使馆遇袭事件“负有责任”，应当是美国无人机“定点清除”的对象。然而，同行的司机和保镖是否都参与了上述攻击事件？是否都对美国国家利益和中东安全造成了“严重威胁”？这些都不得而知。同时，美军在尚未识别全部同行人员身份的情况下，贸然发动攻击，无疑也是违背区分原则、滥用武力的表现。其三，美军造成“附带伤害”的正当性亦存在争议。沃尔泽等正义战争理论家认为，在战区范围内，因为攻击敌方合法的、高价值的目标带来对无辜平民无法避免的“附带伤害”，这在道义上是可以接受的。但是，其前提条件是恐怖袭击事件正在发生或具有迫在眉睫的危急性，唯有如此，平民的生命安全才不得不被置于危险境地，造成的“附带伤害”后果才可以被允许。然而，一方面，刺杀行动的发生地——伊拉克首都巴格达的机场公路并不属于交战区域，另一方面，

① Letter dated 29 January 2020 from the Permanent Representative of the Islamic Republic of Iran to the United Nations addressed to the Secretary – General and the President of the Security Council [Z]. United Nations Digital Library, 2020.

② CROWLEY M, HASSAN F, SCHMITT E. U. S. strike in Iraq kills Qassim Suleiman, commander of Iranian Force [N]. New York Times, 2020 – 01 – 02.

界是一个更安全的地方”。特朗普的讲话及美国击毙恐怖组织头目的事实，确实反映了美国无人机反恐在战术层面取得成功的事实。然而，一种作战手段在战术层面过于成功的表现，也可能导致其在战略层面被过分滥用的后果。特朗普上任以来，对无人机这一反恐利器如上瘾般难以割舍，不断提高无人机“定点清除”的作战频率。特朗普政府大肆宣扬无人机击毙恐怖组织头目的战绩，却刻意回避恐怖主义源头难以根除、恐怖主义势力日益向全球蔓延等根本性、战略性的问题，始终无法实现将世界变成“一个更安全的地方”的愿景，从本质上反映了当前“反恐战争”在目的与手段上的不相称与不对等，美国也将落入“越反越恐”的恶性循环中难以自拔。

四、美国“反恐战争”的非正义性分析

从此次刺杀行动来看，尽管特朗普政府就刺杀苏莱曼尼事件的正当性给出了冠冕堂皇的陈述，但是，在正义战争理论框架下，从开战正义、交战正义的伦理原则出发，对刺杀事件的整个过程进行道义评判，可以充分证明特朗普政府发动和实施此次刺杀行动的非正义性。

第一，在开战正义层面，由特朗普直接策划指挥的刺杀行动不具有满足开战正义要求的正当理由，复仇才是特朗普政府发起刺杀行动的直接动因。无论是针对恐怖主义袭击行使自卫权利，还是针对潜在恐怖主义威胁行使预先防卫权利，都是欺骗、愚弄国际社会和操控国内舆论的拙劣借口，“以眼还眼”“以牙还牙”的复仇心态才是驱使特朗普下决心“清除”苏莱曼尼的直接原因。需要指出的是，特朗普政府确定复仇行动中武力使用限度，并不是依据苏莱曼尼所犯的罪行，而是依据美国希望通过复仇行动收到的效果，由此

时指出：刺杀行动的目的是“阻止伊朗伊斯兰共和国进行或支持对美国或美国利益的进一步攻击，并降低伊朗伊斯兰革命卫队‘圣城旅’部队支持的民兵进行攻击的能力”。[①] 就作战手段与作战目标之间的相称性而言，刺杀行动并未达成其既定的作战目的。相反，刺杀行动进一步刺激了伊朗对美发动攻击行动的意愿和决心。伊朗最高领导人哈梅内伊承诺进行“有力的报复”，而伊拉克民兵武装则高呼“报仇是未来”。伊朗军方随即于1月8日对驻守美国在伊拉克埃尔比勒和阿萨德的两个军事基地进行了导弹袭击。从国际主流媒体对刺杀事件的反应来看，许多媒体评论员撰文发表对美伊关系前景的忧虑，克劳利等人在《纽约时报》撰文指出：这次刺杀行动是鲁莽的单方面升级，可能会带来剧烈而无法预见的后果，并可能在整个过程中造成波及中东地区的影响。[②] 瓦克也在《时代》周刊撰文指出：暗杀进一步打开了中东地区动乱的潘多拉魔盒，伊朗将尝试直接反击美国，并且双方的危机将向潜在的邻国和国际社会蔓延。[③]

从更长远的战略视角来考量，美国使用无人机“定点清除”的作战手段与其在全球范围内彻底消除恐怖主义的战略目标之间也难以相互匹配。刺杀苏莱曼尼事件发生后，特朗普在讲话中炫耀了美国利用无人机手段打击恐怖势力的战果：“在我的领导下，我们摧毁了‘伊斯兰国’的哈里发，最近，美国特种作战部队消灭了被称为巴格达迪的恐怖分子领导人。”他坚定地认为，“没有这些怪物，世

① Letter dated 8 January 2020 from the Permanent Representative of the United States of America to the United Nations addressed to the President of the Security Council [Z]. United Nations Digital Library, 2020.

② CROWLEY M, HASSAN F, SCHMITT E. U. S. strike in Iraq kills Qassim Suleiman, commander of Iranian Force [N]. New York Times, 2020-01-02.

③ VAKIL S. Qasem Soleimani's assassination opens a Pandora's Box in the Middle East [J]. Time, 2020: 1.

界是一个更安全的地方”。特朗普的讲话及美国击毙恐怖组织头目的事实，确实反映了美国无人机反恐在战术层面取得成功的事实。然而，一种作战手段在战术层面过于成功的表现，也可能导致其在战略层面被过分滥用的后果。特朗普上任以来，对无人机这一反恐利器如上瘾般难以割舍，不断提高无人机“定点清除”的作战频率。特朗普政府大肆宣扬无人机击毙恐怖组织头目的战绩，却刻意回避恐怖主义源头难以根除、恐怖主义势力日益向全球蔓延等根本性、战略性的问题，始终无法实现将世界变成“一个更安全的地方”的愿景，从本质上反映了当前“反恐战争”在目的与手段上的不相称与不对等，美国也将落入“越反越恐”的恶性循环中难以自拔。

四、美国“反恐战争”的非正义性分析

从此次刺杀行动来看，尽管特朗普政府就刺杀苏莱曼尼事件的正当性给出了冠冕堂皇的陈述，但是，在正义战争理论框架下，从开战正义、交战正义的伦理原则出发，对刺杀事件的整个过程进行道义评判，可以充分证明特朗普政府发动和实施此次刺杀行动的非正义性。

第一，在开战正义层面，由特朗普直接策划指挥的刺杀行动不具有满足开战正义要求的正当理由，复仇才是特朗普政府发起刺杀行动的直接动因。无论是针对恐怖主义袭击行使自卫权利，还是针对潜在恐怖主义威胁行使预先防卫权利，都是欺骗、愚弄国际社会和操控国内舆论的拙劣借口，“以眼还眼”“以牙还牙”的复仇心态才是驱使特朗普下决心“清除”苏莱曼尼的直接原因。需要指出的是，特朗普政府确定复仇行动中武力使用限度，并不是依据苏莱曼尼所犯的罪行，而是依据美国希望通过复仇行动收到的效果，由此

就可以解释，美国尽管无法就苏莱曼尼策划了袭击行动给出真凭实据，仍然将打击的对象由伊拉克民兵武装扩展到伊朗“圣城旅”指挥官，其行径绝非正当的防卫行为，而是通过刺杀行动威慑和遏制伊朗在中东地区的渗透行动，迟滞乃至终止其发展核武器的进程。然而，实践表明，突破限度的报复行为并不一定能够取得预期的威慑效果，反而可能陷入“犯罪—报复反击”之间的恶性循环，从而加剧暴力传播的程度，并可能使未来美国遭受袭击的风险进一步加大。

第二，在交战正义层面，美国精心打造的无人机反恐技术系统，在一定程度上提升了反恐作战的效率，但也相应地降低了交战过程中使用技术的道德标准。美军使用无人机刺杀苏莱曼尼的交战过程，远非特朗普政府宣称的那样“完美无瑕”。特朗普及其国家安全团队在距离伊拉克万里之遥的美国本土远程指挥，他们听不到战场的声音，也感受不到战场的氛围，这种感官上的隔绝进而导致精神上的割裂，使得他们的反恐作战决策与可能导致的道德后果分割开来。为了达成在第一时间击毙苏莱曼尼的目的，特朗普政府丝毫没有考虑其他同车人员的身份识别和攻击合法性问题，也丝毫没有考虑刺杀行动对于中东地区局势和民众生命财产安全的一系列负面影响。在刺杀事件之后的发言中，无论是特朗普本人，还是美国政府和军方，都没有表露出任何对杀戮无辜者行为的愧疚感。由此可见，虽然美国政府频繁使用“零风险”“零伤亡”“精确打击”“无附带杀伤”“外科手术式打击”等词汇，将无人机作战美化为最具人道主义色彩的反恐手段，但是，理想中的人道主义愿景与现实中的人道主义战争并不同，“反恐战争”本质的残酷性没有改变，这使得两者之间尚存有难以逾越的鸿沟。

第三，新保守主义的道德观念是主导特朗普时代“反恐战争”

逻辑的道义基础。虽然特朗普政府尝试证明其军事行动的正义性，但是，无论给出的理由如何正当、动机如何高尚，本质上都体现了在新保守主义思想驱动下特朗普政府的政治与战略考量，以及借反恐作战输出美式自由民主观念、不择手段地占据“道德高地”等政治诉求。一方面，新保守主义思想赋予特朗普政府实施刺杀行动的道德优越感。在就刺杀苏莱曼尼行动发表的讲话中，特朗普标榜美国“一向追求好人、伟人和伟大灵魂的利益”，将自己塑造为“世界和平、和谐与友谊”的捍卫者，而将苏莱曼尼指控为“恐怖主义军阀”，反映了新保守主义思想所标榜的“民主国家”与“独裁国家”之间“正义与邪恶的区分”，以及“非黑即白”“善恶两分”的道德标准。另一方面，新保守主义思想促使特朗普政府过度追求个体自然权利而忽视共同价值。正义战争理论之所以成为评判战争和非战争武力行动的理论依据，本质上在于它充分契合了人类文明共享的一些基本价值观念，体现了尊重生命、同情弱者的人道主义精神和国际社会普遍认可的原则、共识和法律规范。在刺杀行动实施过程中，特朗普政府的所作所为，公然践踏了正义战争理论和国际人道主义法的基本原则，充分体现了特朗普政府基于所谓民主自由国家天然的道德优越感，片面追求新保守主义思想的自然权利，以及对“独裁国家”“恐怖主义势力”进行执法的自然合法性。

在刺杀苏莱曼尼的军事行动中，特朗普政府凭借无人作战平台及各类信息化装备带来的压倒性技术优势，对主权国家军事领导人肆意实施“定点清除”行动，在全球范围内清除异己，对正义战争理论造成一系列的冲击与挑战。更为值得关注的是，无人机反恐的技术优势如何影响特朗普政府对“反恐战争”的认知、理解并改变其作战实践。特朗普政府将伊朗“伊斯兰革命卫队”“圣城旅”等主权国家军事力量定义为“恐怖组织”，将主权国家军事领导人确定

为反恐作战的目标，进一步扩大了“反恐战争”的打击目标，反映出在技术优势的加持下，特朗普政府和美国军方对“反恐战争”的道德观念正在发生潜移默化的转变。

近年来，美国在反恐情报搜集和远程无人打击方面形成的压倒性优势，不仅强化了其打赢全球化“反恐战争”的手段方式，也相应降低了其发动战争的成本，导致特朗普政府漠视约束战争的道德准则，滥用开战权利。特朗普政府肆意挑战和践踏正义战争的道德底线，主要体现在以下三个方面。

首先，开战理由日益由防卫转变为复仇。美国发动“反恐战争”的初始动因，是为了追捕恐怖分子本·拉登及其所领导的“基地”组织。这一组织在美国本土制造了骇人听闻的“9·11”恐怖主义事件，因此，美国基于防卫恐怖主义袭击的目的，对“9·11”事件的肇事者发起追捕，具有一定的正当性与合法性。伴随智能化无人作战平台、大数据分析工具和技术手段在反恐作战中的深入应用，美国在对恐怖分子的作战中拥有无可撼动的绝对技术优势，并取得反恐作战一次又一次的辉煌战果，“反恐战争”的胜利已成为一个可以自我实现的“神话”。然而，战术手段上的过于成功，也可能导致战略目的上的迷失。无人机让美军实现了非接触、高效率、低成本、零伤亡的“反恐战争”，近乎彻底地清除了发动和持续推进战争的两个制约因素——国内政治成本和军事成本。面对恐怖主义活动在全球范围的蔓延，特朗普政府片面依赖武力“以暴制暴”的复仇心态日益凸显。对于刺杀苏莱曼尼的原因，特朗普在1月13日发表的个人推文中表明心迹：“苏莱曼尼是否计划即将进行袭击并不重要，因为他过去的可怕经历！”这是对特朗普复仇心态的直接表述，也意味着在特朗普政府认同的开战理由中，复仇已经取代防卫成为更具有优先级的理由。而特朗普在白宫的讲话中也指出：“如果任何地方的

美国人受到威胁，我准备并采取任何必要的行动，特别是针对伊朗。"[①] 只要受到"威胁"而非遭受实质性伤害就可以"采取任何必要的行动"，意味着在强大技术优势加持下，美国"反恐战争"日益在战略上走向迷失，复仇心态驱使下的"反恐战争"扩大化趋势已经向全球蔓延开来，从而将国际安全环境拖入更为严峻的困境之中。

其次，"反恐战争"的"合法权威"日益由主权国家的权威转变为总统个人的权威。在经典正义战争论中，对发动战争的合法权威原则有严格界定。合法权威，即战争必须是由代表国家权威的统治者按照公开的、合法的程序施行。合法权威原则的严苛约束条件，原本使得各国政府必须慎重考虑战争决策，而美国无人机反恐的技术优势，使得战争变得毫无风险。美国政府在做出无人机反恐决策时不再谨慎小心，甚至公然违背法律规定的战争程序。在美国以无人机击毙苏莱曼尼之后，国际舆论纷纷抨击刺杀行动的非法性，其中一个争议的焦点，就在于特朗普政府没有遵循合法权威原则，在没有经过国会批准的情况下，以不合乎法律要求的程序实施刺杀行动。2020 年 1 月 13 日，美国众议院以 224:194 的投票通过了一项决议，裁定国会未授权总统对伊动武。[②] 面对外界的质疑，特朗普自信地宣称美国拥有世界上"最好的军事力量"和"世界上最好的情报"，认为"我们已经充分确定了所有的攻击目标"。特朗普政府依仗的技术优势，不仅指无人机强大的打击优势，而且包括支撑无人机反恐的整个情报侦察技术体系。正如《纽约时报》所指出的：

① National Security & Defense. Remarks by President Trump on the killing of Qasem Soleimani [EB/OL]. (2020 - 01 - 04) [2022 - 04 - 07]. https://user.guancha.cn/main/content?id=221304&s=fwzxfbbt.

② GALBRAITH J. U. S. drone strike in Iraq Kills Iranian military leader Qasem Soleiman [J]. American Journal of International Law, 2020, 114 (2): 313 - 323.

“在任何给定的时间跟踪苏莱曼尼的位置，一直是美国和以色列间谍部门和军队的首要任务。现任和前任美国指挥官和情报官员表示，周四晚上的袭击，除却利用了来自线人的高度机密信息，美军还使用了电子拦截、侦察机和其他监视手段。”① 在这一技术体系的支持下，特朗普才敢于在不经过国会授权，甚至不需要公开程序的情况下，授权中情局等机构利用隐秘技术手段远程遥控展开行动。更为值得关注的是，无人机反恐的技术优势，也导致美军的反恐作战决策难以公开透明。在斯坦尼福斯和阿赫加尔等人看来，由于大数据技术在反恐作战决策中的应用，公开的开战程序日益被隐秘的算法、模型和运算程序所取代。② 在无人机“反恐战争”决策过程中，以各类算法、模型为基础的恐怖主义袭击威胁评估系统被视作严格保护的机密。这些不透明、不接受质疑、难以被理解的算法和模型为“反恐战争”决策提供关键依据，国际社会和普通民众无法了解决策程序的具体运行规则，很难对其决策结果的正义性提出反馈和质疑。

最后，战争是“最后手段”的观念日益向“最优手段”转化。美国无人机反恐作战的技术优势，颠覆了逐渐升级的传统危机演进模式。刺杀苏莱曼尼的无人机反恐行动，也反映了特朗普政府对最后手段原则的认知发生了重大变化，“最后手段”日益向“最优手段”转化。在解决国际恐怖主义威胁的过程中，较之暴力程度更低但程序烦琐、时间冗长的政治、外交手段，能够一锤定音且毫无风险的无人机“定点清除”手段，显然更受到政治家的青睐。特朗普政府在没有掌握确凿证据的情况下，就使用无人机击毙伊朗“圣城

① CROWLEY M, HASSAN F, SCHMITT E. U. S. strike in Iraq kills Qassim Suleiman, commander of Iranian Force [N]. New York Times, 2020 - 01 - 02.

② AKHGAR B, SAATHOFF G B, ARABNIA H R, et al. Application of big data for national security [C]. Butterworth - Heinemann, 2015: 23 - 38.

旅”指挥官苏莱曼尼，也反映了无人机作为反恐最优手段推动美国反恐战略向着预防性反恐的方向不断发展。以美国为首的北约国家在“反恐战争”中，广泛应用人工智能技术预判恐怖威胁，通过对过往恐怖袭击时间、空间、人员、装备，以及政治诱因、宗教诱因、民族诱因等历史信息进行关联分析，预测未来可能发生的恐怖袭击事件和可能存在的恐怖分子。据美国国防部官员透露，在刺杀行动之前，美国对于苏莱曼尼的监控和情报搜集已经长达数年之久。[①] 运用大数据情报分析技术，特朗普政府在对苏莱曼尼涉恐情报的分析研判中，可能将遥远的、未知的、难以确定的恐怖主义威胁评估为紧迫的、确定的战争前景，无视正义战争理论对开战正当理由的严苛限制，在国家主权未受侵犯的情况下轻易逾越战争门槛，发动先发制人的“定点清除”式反恐行动。

展望未来，美国无人机反恐作战向智能化发展，可能进一步挑战支撑最后手段原则的均衡性基础。最后手段原则成立的基础是敌对双方力量总体上处于均势状态，从而使他们审慎思考未来战争的残酷性和取得战争胜利的惨重代价。冷战时期美苏双方的核均衡导致超级大国之间必须慎重使用武力，在一定程度上保证了新的世界大战打不起来。美国无人机反恐作战进一步打破了国际军事竞争的均衡性基础。无人化武器系统实现人与武器的实质性分离，极大提升作战效率，同时使己方战争伤亡无限趋近于零，从而在更深层次上挑战了支撑最后手段原则的均衡性基础。

① BROOK T V. Qasem Soleimani: The Pentagon had tracked Iranian general for years before he was killed [EB/OL]. (2020-01-03) [2022-04-07]. https://www.usatoday.com/story/news/politics/2020/01/03/us-military-tracked-iran-general-soleimani-years-killed-thursday/2806630001/.

第七章

智能化战争“慈化”愿景及其实践困境

人类历史上，追求战争的“慈化”是许多道德哲学家、有良知的科学家和技术专家心中共同的追求。早在19世纪，欧洲出现了工业技术和平主义思潮，认为技术的发展能够使战争变得更加道德，并最终能够消灭战争。20世纪90年代的海湾战争，开启了信息化武器装备应用于军事领域的新篇章，“非接触”“低伤亡”“精确杀伤”等成为信息化战争的代言词，引发了人们对于战争日益趋向人道的思考。21世纪初，美国政府使用无人机进行反恐，掀开了智能化战争的序幕，也将“无人化武器装备使得战争变得更为人道”的议题进一步展现在人们面前。智能化战争是否更加符合人道主义的精神？本章将从人类战争武器装备和作战方式的演进过程出发，分析不同时期人类作战手段方式引发的不同伦理问题，揭示由传统武器装备向信息化、智能化武器装备发展过程中战争伦理面临的一系列两难

道德困境。通过宏观考察人类战争历程，进一步揭示少数西方强权国家智能化作战的非正义本质。

一、智能化战争“慈化”愿景

就技术进步的源流而言，智能化战争脱胎于20世纪90年代兴起的信息化战争。无人机作战是信息化战争向智能化战争过渡的一个节点，它既保留了传统信息化战争人类远程控制和精确打击的特点，又展现出智能化分析决策和机器人自主攻击的作战潜能。与之相对应，随着参与战争的人越来越少，人类逐渐退出作战打击平台，机器人逐渐成为战争的主体，战争是否会向“慈化”的方向发展，这一问题引起了人们的深思。

回顾历史，在核武器出现以前，武器的演进长期处于杀伤力“短缺”的时代。从冷兵器到单发火器再到自动火炮，军事技术发展的主线始终体现为对杀伤力的追求。直至第二次世界大战期间原子弹的发明与应用，第一次将“超杀”的概念展现在世人面前，也标志着人类对于纯粹杀伤力的追求走上了穷途末路。海湾战争以来，伴随信息化战争登上人类战争的历史舞台，以精确杀伤为代表的信息化武器装备日益成为战争的主导武器，使得武器进化的主线由追求杀伤转为追求控制，同时也将一幅战争的“慈化”愿景呈现在世人眼前。部分西方学者与军事家乐观地认为，信息化战争既能够精确选择作战目标而不伤及无辜平民，又能够运用恰当的能量对目标给予打击。正如美国空军军法局前副局长邓拉普所言：“近来科学技术的发展燃起了人们的希望：21世纪的战争——如果真的无法避

免，仍然可能以一种更人道的方式进行。”[①] 中国军事理论学者乔良和王湘穗在其著作《超限战》中也提出类似的观点：“精杀（精确）武器和非杀（非致命）武器的出现，是武器发展的转折点，它标志武器第一次不是向‘强化’方向发展，而是呈现出‘慈化’倾向。”[②]

进入21世纪，随着美国在阿富汗战争、伊拉克战争中大量使用无人机进行“定点清除”打击，智能化战争已经成为未来战争发展的重要方向。对此，一些技术乐观主义者（主要是美国的学者）认为，无人化武器装备能够降低战争伤亡、实现精确反恐打击，因而是一种更符合人道主义要求的战争手段。部分美国学者甚至认为，具有完全自主开火权的无人作战平台，可以彻底摆脱人类战争的各种非理性行为，在战场上比人类表现得更加道德，因而能够推动人类战争向更加人道的方向发展。

然而，从军事技术发展的历程来看，在无人化技术给战争带来人道主义改良的表象之下，隐藏的是诸多非人道因素。智能化武器装备广泛应用，虽然规避了一些传统作战手段带来的伦理问题，但也导致了一系列新的伦理困境。无人机“定点清除”行动虽然展现了极高的效率和一系列成功的战例，但却并没有实现“慈化”的战争，也没有从根源上清除恐怖主义势力。与之相反，“9·11”事件以来，恐怖主义势力在全球不断蔓延，已经呈现愈演愈烈的态势。作为“定点清除”行动的反恐利器，无人机沦为恐怖主义炸弹袭击的重要工具，从本质上说明了单从技术维度“以暴制暴”的方式，难以从根本上解决恐怖主义这一困扰世界和平的全球性威胁。

---

① 海沃德. 反暴乱与反恐行动中的空中力量［M］. 高健，任飞，译. 沈阳：辽宁大学出版社，2011：189.

② 乔良，王湘穗. 超限战［M］. 北京：中国社会出版社，2005：67.

二、作战环境存在现实与虚拟的两难困境

智能化战争独特的作战方式，标志着人类战争的战场环境从现实战场向虚拟战场转换，使人类可以规避亲临现实战场的生命风险，但也带来了新的伦理问题。

人类在战争中的道德行为，受到其所处战场环境的潜移默化的影响。在人类经历的战争实践活动中，存在两类截然不同的作战环境。一类战场环境是客观存在的自然空间。伴随人类科学技术的发展，人类战争的自然空间日益由陆地、海洋向天空及外太空不断拓展，人类战争日益成为多维一体的立体化战争。另一类战场环境是智能化战争中由人类创造的虚拟环境。20 世纪中叶以来，现代信息管理系统技术、模拟仿真技术、智能决策技术、虚拟现实技术等各类现代新技术，为人类在虚拟作战环境下指挥作战提供了多元化的工具手段。依托信息管理技术，实现对各类战争信息的采集、传输、综合处理；依托军事仿真技术，对军事系统内部要素进行建模，实现武器技术仿真、武器系统仿真及作战系统仿真；依托智能化的决策支持技术，实现战场情报融合、打击目标分析、作战方案仿真、作战计划运筹分析；依托虚拟现实技术，生成多维战场感知，实现与虚拟战争环境的交互作用，提供身临其境的真实感。

对于区分原则而言，无论是客观存在的自然空间，还是人类创造的虚拟环境，都存在着难以规避的伦理困境。在冷兵器和热兵器时代，人类在传统自然空间主要依靠肉眼进行区分，区分的结果受到武器作战半径不断拉伸的影响，造成遵循区分原则的困境。然而，在信息化战争以及当前智能化无人作战中，虽然武器平台运行和火力打击的过程仍然在自然空间展开，但是，作战人员日益从武器平

台中脱离出来，在距离现实战场万里之遥的指挥中心，通过视频设备呈现的虚拟战场图像来实现远程遥控作战，由此产生与传统战争截然不同的作战体验，进而影响作战人员的作战区分行为，导致新的区分困境。

（一）现实环境下区分原则的弱化

长期以来，体现作战目标准确选取的区分原则，一直是构成战争伦理理论体系的重要原则。“区分原则强调武力使用过程中应在平民与战斗人员之间和民用设施与军用设施之间加以区别，武力应仅以战斗人员和军事设施为目标，而绝不能有意以平民和民用设施为目标。”① 作为人道主义诉求在军事领域的集中体现，对军事目标与非军事目标的区分对待，成为古今战争史上有良知的军事家和政治家所认同的基本道德规范。从冷兵器时代到20世纪中期，人们对于军事目标与非军事目标的识别和打击，一直是在真实的战场环境下进行的。在这一漫长的时域内，一方面，武器的攻击距离与作战半径持续增长，特别是19世纪末20世纪初空基作战力量的出现，使得这一增长急剧加速。另一方面，对作战目标的识别与区分能力却没有得到同步的增长，长期处于依靠原始感官系统的低水平阶段。如此一来，随着武器的攻击距离超越人类的视距，作战人员对于军事目标与非军事目标的区分能力越来越弱。同时，由于作战人员看不到敌军和平民被攻击后的情形，他们对于杀戮行为的愧疚感也随距离的增长而逐渐消减。对于现实环境中因目标区分困境而引发的道德问题，美军退役中校格罗斯曼认为，当物理距离愈远时杀人愈轻松，参战人员抵触杀戮的意愿与其至目标的距离之间呈现反比关

① 朱之江. 现代战争伦理研究［M］. 北京：国防大学出版社，2002：156.

系。区分性原则的逐渐弱化，导致战争的附带伤害状况持续恶化。在第一次世界大战和第二次世界大战中，生物武器、化学武器及核武器等大规模杀伤性武器的问世和应用，造成了对大量无辜平民不加区分的攻击，酿成了许多战争惨案。这一时期军事技术的发展，造成了战争与人道精神之间更大程度的背离。

（二）虚拟环境下选择性杀伤的回归

在冷兵器时代，由于武器攻击范围和攻击强度的局限，因而可以在军事行动中实现对目标的选择性杀伤。然而，从火器发明后的热兵器时代开始，一方面军事技术发展始终以追求杀伤的最大化为主线，另一方面军民界限日益模糊，且火器所具有的随机性并非选择性的杀伤能力，致使军队在施行军事行动时必然拥有无情的实用主义精神。

精确制导技术在信息化战争中的广泛应用，从技术层面为选择性杀伤在军事领域的回归提供了可能。具体而言，精确打击武器为现代战争创造了一幅全新的画卷，精确与清白成为这一画卷的代名词。军事理论家们常在精确打击前面冠之以“外科手术式”的修饰词，以技术隐喻的形式将两者联系在一起。在理想化的“清白战争”中，作战人员扮演外科医生的角色，精准地操作手术刀，切除感染或病变的组织并干净准确地缝合伤口，而不造成任何的附带伤害。情况正如哈佛人权政策中心主席伊格纳蒂夫所言：“我们将自己看作高贵的战士，将我们的敌人看作可耻的暴君。我们将战争看作一把外科手术刀，而不是沾满血污的利剑。”[①] 随着人工智能技术军事应

① 海沃德. 反暴乱与反恐行动中的空中力量［M］. 高健，任飞，译. 沈阳：辽宁大学出版社，2011：202.

用程度的不断深化，具有自动寻找、识别、跟踪和摧毁目标功能的无人机、无人驾驶坦克、无人操控火炮、智能鱼雷和自主多用途智能作战机器人，开始覆盖战场的每个角落。目标识别能力的大幅度提升和各类智能化武器装备的使用，使得军事对抗向着准确区分作战目标的人道主义方向迈进了重要一步，这种防止产生附带伤害的选择性杀伤，无疑体现了“慈化”战争的道德愿景。

（三）物理距离的缩小与心理距离的拉大

在现实战场环境下，作战目标的区分存在距离与道德的问题，与距离的增加相伴的是作战人员杀戮意愿的增强，以及对杀戮行为的冷漠与麻木。

在智能化作战模式下，目标感知和识别技术的发展，使得对于作战目标的区分由现实环境转为虚拟环境。在美国无人机反恐作战中，反恐作战人员借助卫星实时传输视频信号，可以克服作战目标超视距的困境，实现对恐怖分子的识别和选择性杀伤。考虑到电磁网络空间的物理属性，无人机反恐无疑在物理层面缩短了视觉距离，实现对恐怖分子的精确选取，从而在最大程度上避免无辜平民的伤亡。无人机反恐在虚拟环境下极大地缩短了识别目标的物理距离，从道德和距离的相关性来看，这一技术上的进步无疑与战争的“慈化”愿景相吻合。

然而，值得警示的是，智能化战争中的虚拟战场环境，缩短了作战人员与作战目标之间的物理视距，同时也拉大了两者之间的精神距离。在虚拟环境下，反恐作战人员通过显示器屏幕来实现对恐怖分子的识别与区分，并借此实现对目标的打击，但他们完全听不到战场声音，也嗅不到战场的气味。这种感官上与现实战场的隔绝，可能会逐渐演化成精神上的割裂，武器的操作人员日益与他们的行

为所产生的后果分割开来，其结果是抑制了作战人员对杀戮行为的愧疚感，进而威胁到他们作为军人的专业道德和作为人类的基本良知。一份关于无人机“定点清除”行动的联合国官方报告指出，作战人员对于无人机的控制“完全可以通过计算机屏幕和远程声频反馈来实施，因此存在以游戏心态来看待杀人的风险”。[①] 无人机的操作者以游戏的心态应对流动的战场和多变的情报，无疑会增加目标选取的失误率，攻击平民的错误杀戮行径也就会持续出现。“9·11”事件以来，美军在世界各地的“定点清除”行动中频繁发生的袭击平民事件，无疑印证了联合国官方报告中的结论。

三、作战手段运用存在滥杀与精杀的两难困境

长久以来，人们将对生命的剥夺看作战争暴力本质的重要体现，也是战争与人道主义原则的根本冲突所在。在古代和近代战争中，由于缺乏精确杀伤武器，战争往往演变为灭绝人性的大屠杀，无疑是与人道主义精神背道而驰的。有鉴于此，依靠科学技术的进步，谋求不危及或少危及人类生命的作战手段，无疑是合乎人道主义原则的理性选择，也成为智能化战争中彰显人道主义的重要主题。然而，当我们用理性的眼光去审视现代战争中独领风骚的各类精确杀伤武器，就会发现所谓的精确杀伤与真正意义上的人道主义精神之间仍然存在难以逾越的鸿沟。

（一）致命手段的持续强化对相称性原则的挑战

战争伦理是在人类长期战争实践中逐渐形成的，而战争伦理的

① Report of special rapporteur on extrajudicial, summary or arbitrary executions [R]. General Assembly, 2010: 25.

调节作用主要是通过敌我双方在战争实践中的复杂对抗手段体现出来的。因此，围绕战争伦理的探讨，也主要体现为如何规制战争手段，使其得以合理地运用。具体而言，战争手段的运用应当符合相称性原则，即允许使用的作战手段所造成的损伤程度大致与预期取得的直接军事利益成比例，尽量规避不合理使用战争手段而带来的不必要的附带伤害。

在原子弹发明以前，人类战争处于致命手段持续强化的阶段，利用科学技术成果，研制具有更大毁伤威力的致命装备，成为这一阶段作战手段发展的主线。在20世纪上半叶的两次世界大战中，致命手段的持续强化达到了登峰造极的地步。各主要参战国遵循总体战的战略思想展开全面的战争动员，从而将国家的政治、经济、科技、军事和文化力量集中投入战争实践中，并研制出一系列威力惊人的杀人武器，这必然造成作战手段的运用摆脱相称性原则的约束。一战中化学武器的大规模应用以及“二战”中生物武器与核武器的应用，都是违背相称性原则，无所不用其极地强化致命手段的典型案例。

（二）精确杀伤手段实现毁伤的可控性

机械化战争向信息化战争、智能化战争的转变，意味着武器装备的发展由追求最大额度的能量转换与释放，转向追求最精确的能量操控。精确杀伤武器的出现与应用，既不会对目标造成永久损伤，又可以在一定时间内使目标丧失作战能力，从而在作战手段层面将人类对于能量的控制展现到极致。在精确制导技术的支撑下，美国无人机能在保全己方参战人员和平民安全的同时，以最精准、最恰当的手段方式对恐怖分子给予杀伤，使得充分遵循相称性原则的军事打击成为可能。情况正如乔良和王湘穗所言：“慈化武器是人类在

武器领域进行多种选择中最新的一种自觉的选择，它在给武器注入了新技术的因素之后，又加入了人性成分，从而破天荒地使战争涂上了温情色彩。”① 作战手段的精确操控，使得能量杀伤效果处于可控制的范围内，这正是慈化愿景的勾绘者们所力图实现的重要主题。以精确制导武器为代表的信息化作战手段，通过在人员杀伤机理和装备毁伤机理方面的改变，减少了作战中使用大规模杀伤手段的频率和范围，使得毁伤的可控性得以进一步加强。

（三）显性的人道与隐性的非人道

在当前信息化向智能化迈进的时代，人类操控杀伤的能力进一步提升，美国无人机反恐利用“地狱火”精确制导炸弹实现“定点清除”式打击，即是对人道主义诉求的一种回应。较之传统战争，无人机主导的“定点清除”式反恐不需要己方作战人员直接参与，作战范围精确限定，作战杀伤少，战争污染小，对基础设施和环境的附带破坏减低。较之过往武器装备带来的“硬摧毁”“地毯式饱和轰炸”等剧烈杀伤方式手段，“定点清除”式打击对于战场的破坏无疑是温和的，并在某种程度上与人道主义愿景中珍爱生命的主题相吻合。

然而，在显性的人道主义关怀背后，智能化战争还隐藏着隐性的非人道主义因素。美国无人机反恐作战的实践表明，作战手段由滥杀向精杀的转变，并不意味着人道的作战手段取代了过往战争中不人道的作战手段。

为应对恐怖主义势力在全球范围内的肆虐，无人机因其特殊的武器作用机理，成为在人口密集地区遂行反恐任务时最可接受的武

① 乔良，王湘穗．超限战［M］．北京：中国社会出版社，2005：68．

器。然而，这类武器所谓精确的“定点清除”，使得反恐作战人员在面对平民和民用目标时不再谨慎小心。出于这一心态作祟，奥巴马和特朗普两任总统在任期内不断增加无人机在反恐作战中的使用频度，将其滥用于军民目标混杂的反恐军事行动之中，从而造成对区分原则的进一步弱化。此外，“定点清除”式打击在各类反恐行动中广泛应用，造成大量家庭破碎和“从天而降”的人间惨剧，其造成的物理毁伤虽小，但对于受害者的家人和普通民众而言，可能造成强烈的精神伤害和心理恐慌，且这种心理层面的伤害难以准确衡量与评估。对于以家族和部落为基础的伊斯兰社会而言，大量失去亲人的家庭和宗教族群，无疑会将这种心理恐慌快速传递到社会各个阶层，进而可能导致各类大规模群体性事件的发生。因此，较之无人机“定点清除”所带来的人道性的一面，它所具有的非人道主义因素对战争伦理的挑战，更值得人们的关注与反思。

四、作战结果预测存在未知与已知的两难困境

智能化战争区别于以往战争的一个重要特征，就在于利用大数据、人工智能技术等手段方法，对于战争的结果进行预先判断。20世纪中叶以来，系统科学、信息科学、控制科学的蓬勃发展，为人类预测战争的结果提供坚实理论基础。系统科学的发展，为整体把握战争复杂系统提供方法论依据；信息科学的勃兴，为采集、处理与分析设计战争所需信息提供科学方法；控制科学的成熟，为推演和预测战争提供控制机制和方法。得益于科学技术的推动作用，“战争从实验室打响”，已经从军事家的理想转变为性命攸关的现实。构建信息化条件下的作战实验室，综合运用以计算机为核心的各类信息技术手段，模拟可控可测的战争环境，推演敌我双方在多维空间

的军事力量运用，对作战行动进行预先设计，实现军事决策的最优化和作战效果的最佳化，已经成为世界军事强国在开战之前的通行做法。

从战争伦理的视角来看，已知或未知的战争结果会对人类的战争进程造成哪些不同的影响？进而会引发哪些不同的伦理问题？这些问题值得进行深入的探讨。

（一）未知情况下的非理性抉择

古往今来，人类对于信息的掌控与战争的胜负有着密切的关联。《孙子兵法》提出的“知彼知己，百战不殆”，正是基于对战场信息的完全掌控。信息充分完备，就可以精确预测战争的胜负；反之，信息不充分甚至完全缺失，就只能见招拆招，摸着石头过河，甚至面临失败的深渊而浑然不知。战争行为由于关乎民族、国家生死存亡，自古为国之大事，因而对战争胜负的预测自然成为国之要务。尉缭子所说的“战不必胜，不可以言战；攻不必拔，不可以言攻”①，说明只有准确预判战争的结果，才能够做出理性的作战决策。

事实上，古今中外的发动战争者，无一例外地希望获取战争的胜利。然而，在信息技术并不发达的古代，人们对于战场信息的掌控并不完备，“知彼知己，百战不殆”的制胜法则只停留在理想层面。在克劳塞维茨等近代军事家眼中，战争是一个充满偶然性和不确定性的领域。处于战争迷雾笼罩下的战争中，人们无法准确预知战争的胜负，在结果未知的情况下，军事家基于对战争形势的误判，往往会做出非理性的抉择。在中外军事史上，先后出现的曹操进攻

① 尉缭子. 尉缭子兵法［M］. 太原：山西古籍出版社，2003：109.

东吴、拿破仑进攻俄国以及希特勒攻打苏联等事件，都是军事家在信息不完备的情况下做出非理性决策的典型案例。

（二）已知情况下战争的理性设计

20 世纪下半叶以来，科学技术领域的一系列发明与突破，使各国军事家看到了冲破战争迷雾的历史契机。在科学理论方面，涌现出系统论、信息论、控制论和博弈论等理论体系；在技术方面，虚拟现实技术与仿真技术等综合性技术体系逐渐形成与发展；在应用方面，联合作战实验室的建立成为联结作战构想与军事实践的重要环节。利用上述科技领域的最新进展，推演战争的进程并预测战争的结果成为信息化条件下局部战争的重要特征。早在海湾战争期间，美军就利用军团作战模拟系统，对战争进程进行仿真实验和模拟推演，提出名为“100 小时战争”的作战方案，成为后来实施“左勾拳”作战行动的重要依据。近几场局部战争的案例似乎向人们证明，通过仿真实验的方法对复杂的军事系统进行模拟，可以精确地预言和勾绘实战，可谓是战争尚未开始，胜负已在预计之中。情况正如中科院王飞跃研究员所言：“构建各类完备可计算的军事‘基函数’：‘基武器’、‘基靶场’、‘基战场’、‘基军演’、‘基作战’等等，构成虚实互动的平行军事体系，使平行军事分析、实验、训练、作战、评估等成为常规常态的军事活动。这一思路，不但有可能为跨域军事行动提供一个可操作、可计算、可验证的体系，更希望能够形成‘平行军事事务’，开启并引导新的世界范围的军事事务革命。”[①] 事实上，近年来美军一再刷新现代战争的最低伤亡记录，也

① 王飞跃. 面向赛博空间的战争组织与行动：关于平行军事体系的讨论 [J]. 军事运筹与系统工程，2012 (9)：7.

与其对战争结果的准确预测密切关联，由此似乎可以推断，如果基于精确仿真科技的平行军事体系真正实现，不仅会给战争本身带来革命性的变化，也可能为实现对战争的人道主义改良提供新的契机。就伦理层面而言，准确预测战争的结果，实现战争的理性设计，降低决策失误带来的风险，避免重复战斗的需求，进一步降低作战人员的伤亡，无疑是符合战争的道义和人道主义准则的。

（三）均衡性的缺失与战争门槛的降低

在当今智能化战争中，也存在对战争进行理性设计的环节。出于规避战争风险与代价的考量，美国中情局反恐中心利用大数据分析、作战推演、模拟仿真和效果评估，对未知的战争结果进行推演和预判，不仅可以避免无谓的伤亡，而且可以增加成功的概率，降低进行持续消耗作战的可能。然而，对于未知战争结果的恐惧，恰恰是使得许多潜在的战争得以避免的重要原因。从伦理上讲，战争应当受到均衡性原则的约束，敌对双方维持均势状态，会使得他们审慎思考未来战争的残酷性和取得战争胜利的惨重代价。正如美国内战期间著名将领罗伯特·李将军所言：战争残酷是件好事，否则人们真的会爱上战争。因此，一旦战争的结果从无法预知转变为可以预知，并且这一结果正是政治家和军事家所期待的，那么他们就可以轻易跨越战争门槛肆意地滥用武力。

伴随人工智能技术的军事应用，战争的均衡性原则被打破。由于作战结果由未知变为已知，智能化打击方式在西方政治家眼中已经成为一件几乎无风险的机器，战争的胜利也就成为一个可以自我实现的预言。美国使用无人机进行“定点清除”反恐打击，不仅可以规避大规模“反恐战争”造成的人员损失风险，而且可以降低发动“反恐战争”的政治成本，做到迅速、高效、胜利地结束“反恐

战争”。正如黑斯廷斯所述，通过无人机执行遥控任务，五角大楼和中央情报局现在不必向地面派出一兵一卒，就可以实施军事打击或暗杀，也不用担心阵亡美军士兵运回国内而引发的舆论反弹。无人机的隐秘性使得国家领导人能够比以前更方便地发挥美国的军事力量，也使得这些秘密攻击的后果比以前更难评估。隐蔽的、低成本的无人机反恐作战和作战结果的可预期性，不仅使得美国军方和中情局可以轻易逾越发动“反恐战争”的界限，而且可以肆意降低发动战争的门槛。在这一过程中，他们以对“反恐战争”胜利的预期为凭据，以规模日益庞大的无人机编队来替代地面部队，用“零风险”“零伤亡”“无附带杀伤”“外科手术式打击”“首发制穴”等词汇，为自己打造一面“道德盾牌”，借此缓解来自国际和国内的道德压力，一次又一次地开动“反恐战争”机器。由此可见，依靠科技优势来实现战争的人道主义愿景，已经背离了人道主义的道德主题，成为西方政治家缓解舆论压力和实施战争动员的工具，从而助长了他们在战争与和平的抉择中不断挑战更低的道德界限。

综上所述，在人类战争样式由信息化战争向智能化战争的演进过程中，以智能化作战平台、精确打击武器为支撑的美国“定点清除”式反恐作战，仍然难以摆脱人类战争固有的伦理困境。所谓高技术条件下战争的“慈化”愿景，只是某些政治家、军事家和学者基于各自特定目的而创设的概念。“慈化”理念的呈现与“慈化”武器的应用，并不意味着战争已向仁慈方向实现了重大转变，也并不能掩盖战争在伦理层面违背人道主义精神的地方。即便是作为“慈化”战争概念的较早提出者，乔良和王湘穗也不得不承认：“战争乃死生之地，存亡之道，容不得丝毫的天真。哪怕有一天所有的武器都被充分慈化了，一场或许可以避免流血的慈化战争却依旧是战争，它可以改变战争的残酷过程，却无法改变战争强制性的本质，

因而也就无法改变残酷的结局。”[①] 由此可见，理想中的“慈化”愿景与现实中的人道主义战争并不等同，战争本质的残酷性没有改变，这使得两者之间距离仍然遥不可及。

① 乔良，王湘穗. 超限战［M］. 北京：中国社会出版社，2005：69.

第八章
智能化战争风险的规制进路

马克思主义战争观是我们认识现代战争的根本指南。面对全球范围内如火如荼的人工智能军事化发展态势，以及由此引发的一系列战争伦理风险，只有坚持马克思主义战争观这把分析战争问题的金钥匙，才能够透过纷繁复杂的表象，牢牢抓住智能化战争引发伦理风险的本质原因。

从经济维度来看，马克思主义战争观认为，在阶级社会，私有制是战争的根源。智能化战争引发的伦理危机，从本质上体现为经济全球化时代以美国为代表的垄断资本主义国家试图利用人工智能领域的技术优势，强化对发展中国家的经济剥削，从而进一步激化了发达国家与发展中国家之间的矛盾与冲突。从政治维度来看，马克思主义战争观认为，战争是政治的延续。由智能化战争引发的伦理危机也可以解读为美国强权政治和霸权主义在战争领域的延续。

他们利用智能化武器装备是为了进一步巩固其全球政治霸权，实施全球“反恐战争”的本质是为霸权政治张目。从军事维度来看，马克思主义战争观认为，人是战争的决定性因素，也是战争的唯一主体。武器是战争的重要因素，但不是决定的因素，决定的因素是人不是物。当前智能化战争引发的一系列人道主义灾难，从表面上看是武器装备和作战手段的滥用，其背后的根源是人的价值观的扭曲，是珍视生命的人道主义精神受到挑战和威胁。

从马克思主义战争观的视阈出发，可以清醒地认识到，智能化战争并不是某些政客、军方人士所吹嘘的正义的、人道的战争。相反，智能化战争引发的伦理风险、安全风险及人道主义危机已经成为一个全球性问题。如何对智能化战争的反人道主义行径进行规制，避免未来作战中滥用智能化武器装备导致大规模人道主义灾难？

一方面，需要科技领域、哲学领域、法学领域的专家学者集思广益，从技术、哲学、法律等不同路径出发，就智能化战争的规制进路提出专业性建议。另一方面，需要国际社会共同努力，采取积极有效的措施应对智能化战争的全球性风险。正如习近平主席在致信祝贺2018世界人工智能大会开幕时强调，处理好人工智能在法律、安全、就业、道德伦理和政府治理等方面提出的新课题，需要各国深化合作、共同探讨。[①] 基于以上两个方面的考虑，本章主要从技术、哲学、法律、军控和舆论等方面探讨智能化战争风险的规制进路。

① 习近平. 共同推动人工智能造福人类［N］. 人民日报（海外版），2018－09－18（1）.

一、技术规制进路

就技术进路而言，要使得智能化战争遵循战争伦理规范，科技专家主要是探索对智能化战争控制系统实现有意识的技术设计，依靠计算机编程的方式，使之根据一组固定的道德规则运行，从而符合战争伦理和战争法的规定。如阿金认为，我们可以通过给自主武器编程，使之永不违反战争法。[①] 董青岭指出，技术规制不应局限于宏观伦理原则的程序化，还应当涵盖作战应用的具体行为准则。他提出，应当通过诸如导航定位技术预设的方式，禁止自主武器系统在人群聚集或人员流动频繁的城镇执行作战任务，且应限定其执行情报搜集、监视和救援及具有防御性质的特定军事任务，避免造成误判、误伤等安全问题。[②]

虽然自上而下的编程能够构成智能化作战必须遵循的一套规则，但在作战应用的过程中也存在一系列难以规避的问题。卡尔皮等人指出："这些自上而下预置的道德体系缺乏必要的灵活性，无法适应在不断变化的环境中做出正确的道德判断。"[③] "区分盲区"和"跨文化障碍"，以及进行相称性原则道德评估的片面性倾向，都是这种技术进路缺乏灵活性的具体体现。

考虑到依托算法的交战规约预置在作战实践中缺乏必要的灵活性，特别是针对自上而下的编程式道德预置可能存在的伦理风险，

---

① 沙瑞尔. 无人军队：自主武器与未来战争［M］. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：320.

② 董青岭. 新战争伦理：规范和约束致命性自主武器系统［J］. 国际观察，2018（4）：63.

③ KARPPI T，BHLEN M，GRANATA Y. Killer Robots as cultural techniques［J］. International Journal of Cultural Studies，2018，21（2）：107－123.

人工智能专家提出将自上而下的道德预置与自下而上的道德生成相互结合的技术进路。在利用自上而下的编程预先设置基本规则的基础上，通过真实战场环境“改进—继续—改进”的机器学习过程来不断修正既有算法偏见，实现区分目标和相称性评估能力的持续增长。这种自下而上的交战规约生成，在一定程度上可以作为自上而下预置模式的有益补充，实现机器人自主学习和人类能动干预的双向互动、动态发展，帮助无人作战平台在战场的所作所为符合人道主义精神和战争规约要求。但是，在实现区分原则要求方面仍然存在一定的伦理风险。一是机器学习的样本选择偏差，可能导致智能化战争操作系统的错误学习，导致其背离人类程序员最初设定的算法和战争的道义原则。二是机器学习系统运行的透明度问题。自下而上的机器学习系统在生成结果时是不透明的，人类难以理解机器学习或基于规则的算法运算与修正如何得出正确的区分结果。三是机器学习系统的复杂性问题。系统设计缺陷、软硬件故障、操作错误、意外环境交互和黑客行为等，都可能导致反恐无人机发生不可预测的区分风险。

此外，对智能化武器装备的技术设计还必须贯彻必要的透明性原则。增加无人作战平台设计过程中的透明度，对于评判智能化战争作战平台战场行为的合规性，以及精准追究战争责任人具有积极的意义。具体而言，在内在设计中确保必要的透明性主要体现在以下三个方面。首先是可理解的透明性。无人作战控制系统的设计过程应当规避人工智能“技术黑箱”带来的认知误区，充分掌握设计过程中的算法规则、运算过程及后续实现，深入理解每项操控实践可能导致的伦理、法理后果。其次是可交流的透明性。承担智能化作战平台研发核心任务的算法工程师、数据工程师需要充分地与团队内部成员分享设计理念、数据采集，以及装备可用性、安全性、

可扩展性情况，对无人作战平台在装备测试、演习演训中出现的问题展开集体研讨，规避样本的单一性和算法的偏见性，最大限度实现数据采集的多样性、客观性及算法设计的无缺陷性。最后是可公开的透明性。在遵守军事保密条例的前提下，各类智能化作战平台的设计、使用应该适当公开人工智能算法的操作规则、运算过程、后续实现及验证目标的相关记录，实现技术事故源头的可追溯性。

二、哲学规制进路

哲学进路主要关注隐含在技术设计背后的价值观念。伦理价值观念不仅影响人类的行为，也可以在智能化武器装备的技术设计过程中得到充分显示。智能化武器装备作战应用引发的伦理问题，在很大程度上是由人工智能技术军事应用造成的。按照温纳等技术哲学家的看法，技术的本质并非作为一种价值中立工具，价值观念总是嵌入技术设计中。乌姆布雷洛指出，对智能化武器装备的伦理规制必须充分考虑其所隐含的价值观念。卡洛等人进一步指出："如果没有认识并解开嵌入到框架本身设计中的价值观，更不用说那些嵌入在所涉及的机器人技术中的价值观，可能会导致形成错误的固定性思维模式，从而错误地将不合理、致命性的操作视为军事需要。"① 由此可见，如果不能充分审视价值观念的合理性与局限性，那么对于智能化武器装备的技术设计就会误入歧途，甚至带来巨大的伦理风险，导致人道主义灾难的发生，这也是从哲学进路规制无人机反恐作战的必要性所在。

① 卡洛，弗鲁姆金，克尔. 人工智能与法律的对话［M］. 陈吉栋，董惠敏，杭颖颖，译. 上海：上海人民出版社，2018：359.

有鉴于此，未来对于智能化武器装备控制系统的设计研发过程中，必须充分考虑隐含的伦理价值观念，将正义战争所代表的人道主义价值观预先嵌入基础算法和程序的设计之中。鉴于人工智能技术军事应用的复杂性及其重大社会影响，深入研究预先嵌入哪些伦理规约、如何在技术设计中嵌入伦理规约等问题，应当成为伦理学界和人工智能领域学者合作探索的方向。

具体而言，对于智能化武器装备的不同技术设计方案，反映的价值观念不尽相同，对战争和社会发展造成的伦理后果也不尽相同。例如，阿金提出以计算机编程方式来使智能化武器装备在作战过程中遵守战争规约的技术路线，其背后所反映的是一种功利主义主导的伦理价值观念，即只关注智能化武器装备作战结果的正义性，而忽视决策过程的正义性。针对这种功利主义的价值观，莱弗林豪斯指出，即使构想一个理想的场景，智能化武器装备在战场上表现完美，不会违反任何战争规则，但智能化武器装备在战场上的滥用仍然代表了战争伦理规范上不受欢迎的转变。[①] 大数据分析决策过程中“技术黑箱”的存在，导致人类正在冒着丧失尊严和人权的危险。这种危险类似于马克思所指出的“机器对人的反向统治”，人工智能系统的创造意味着权力的根本转变，其中智能化武器装备不再是人类用于特定目的的战争工具。相反，具备高度智能化的无人机控制系统有自己的目的和判断，人类的角色是监督和防范中断。在这种关系中，人类的活动是由智能系统的运动决定和调节的，而不是相反的。

功利主义价值观主导的技术规制可能存在一系列伦理风险。当

① LEVERINGHAUS A. What's so bad about killer robots? [J]. Journal of Applied Philosophy, 2018 (2): 341 - 358.

前，学界普遍认为，人类对智能化武器装备的设计嵌入价值观，首先应当考虑智能化武器装备可能出现的失控情况，因而必须坚持将智能化武器装备牢牢掌握在自己可控的范围内。阿萨罗提出针对人工智能系统进行有意义的人类控制，即为了确保智能化武器装备始终由人类进行控制，任何作战决策都应该由人类自己独立做出，从而维护体现人类美德的人道主义原则，并且不会在此过程中使人类变得无足轻重。[①] 然而，什么样的控制是有意义的控制？人类如何对智能化武器装备进行有意义的控制？学界尚存在一定争议。在笔者看来，有意义的控制至少应当体现在四个方面。

首先是可理解的控制。即人类对智能化武器装备的技术设计应当规避人工智能“技术黑箱”带来的认知误区，充分理解设计过程中的算法规则、运算过程及可能导致的道德后果，特别是理解剥夺他人生存权的残酷性和道德意义。

其次是可干预的控制。智能化武器装备投入作战不能以侵犯人类控制武器的自主性为代价。人类对于无人机的控制，不应是被大数据反恐研判牵引的、亦步亦趋的被动控制，而应是基于人类道德逻辑的主动干预；人类应当将战争决策和战争暴力使用的控制权掌握在自己手中，而不应该过度依赖人工智能技术，甚至被其反向操纵或胁迫。人类的控制应当作为最后一道“保险”。一旦智能化武器装备存在失控的风险，人类就要有足够的能力、充裕的时间终止程序，避免发生大规模人道主义灾难；一旦智能化武器装备的运行存在违背人道主义的潜在风险，人类就要迅速干预以纠正或停止其操作。

---

① 卡洛，弗鲁姆金，克尔. 人工智能与法律的对话［M］. 陈吉栋，董惠敏，杭颖颖，译. 上海：上海人民出版社，2018：384.

再次是可追责的控制。人类不能局限于激活智能化武器装备的角色定位，而应当将自己置于人道主义原则和战争法约束下的控制者角色，为控制行为承担应有的责任。

最后是全过程的动态控制。人类的控制应当遵循“人在回路”的主动监管原则，覆盖智能化武器装备系统从设计、研发到应用等每一环节，同时，应让控制措施适应日新月异的技术发展和不断变换的战场环境。

三、法律规制进路

法律进路主要是探讨如何依法治理智能化武器装备的滥用现象，以及如何破解智能化武器装备作战的“责任鸿沟”，实现对其依法追责。

对滥用无人机反恐的法律规制，集中体现在制定智能化武器装备的法律规约。就国家层面而言，为避免未来作战中滥用智能化武器装备导致大规模人道主义灾难，为杜绝完全自主武器系统造成的“机器暴政”，为规避人类沦为机器猎杀目标的恐怖景象，必须制定相关法律制度，使得智能化武器装备的部署和使用过程有法可依，相关法规核心要义应当包括三方面。首先，智能化武器装备使用的目的限于防卫而非进攻。严格意义上，防卫是正义战争的唯一正当理由。智能化武器装备的部署与使用应当仅限定于执行纯粹防卫性质的军事行动，诸如对恐怖主义活动的侦察监视、情报搜集、危机预警和应急救援等非攻击性军事任务，而不宜将智能化武器装备使用在大规模战争或易于造成误判、误伤的“定点清除”或“斩首”行动之中。其次，智能化武器装备使用范围限于非平民聚居的交战区域。鉴于当前智能化武器装备无法准确区分军人与平民，因而智

能化武器装备的使用范围应当严格限定于沙漠、荒原、深海等无人区域，禁止在人类活动频繁的区域特别是人口密集的生活区使用无人机进行反恐作战。最后，智能化武器装备火力打击的使用程度限于非致命。对生命权利的剥夺是战争与人道主义原则的根本冲突。谋求不危及人类生命的作战手段，无疑是合乎人道主义原则的理性选择。因此，尽可能在对恐怖分子嫌疑人的定点打击中使用非致命性武器，可以在一定程度上防止反恐滥杀行为引发的人道主义灾难，也可作为人道主义精神在反恐作战中的重要体现。

此外，法律规制路径还包括构建和完善智能化武器装备作战失范行为的责任追究机制。立足当前人工智能技术的发展水平，国际法学界人士普遍认为，破解智能化武器装备作战应用过程中的责任缺位难题，关键在于厘清背后相关责任人的法律责任。对于设计开发人员的责任追究，阿萨罗认为，智能化武器装备的设计开发人员应当对编程错误或硬件设计失误引发的战争罪行负责。一方面，如果设计开发人员故意设计具有编程缺陷、造成操作故障或没有道德约束感的反恐无人机，他们必须为由此产生的人道主义灾难负责；另一方面，如果是无意识的疏忽，开发人员同样需要为未能提出警告或未能采取适当谨慎措施而承担法律责任。[①] 对于军事指挥官的责任追究，舒尔茨克认为，指挥官有责任确定智能化武器装备将被投入作战行动的条件，如应当避免其在平民聚集区域作战，并确保任何被批准在智能化战争中服役的操控员已经过彻底筛选。[②]

立足未来人工智能技术不断发展的趋势，也有一些学者认为智

① 卡洛，弗鲁姆金，克尔. 人工智能与法律的对话［M］. 陈吉栋，董惠敏，杭颖颖，译. 上海：上海人民出版社，2018：383－384.

② SCHULZKE M. Autonomous weapons and distributed responsibility［J］. Philosophy & Technology，2013（26）：203－219.

能化武器装备的“责任鸿沟”将不复存在。苏林斯认为，当机器人的智能化水平足以支撑其进行符合道德规范的自主行动时，它们将成为道德行为者，将享有与所有道德行为者相同的权利和责任。[①] 然而，即便未来智能化武器装备成为真正意义上的道德行为者，也并不意味着能够对其实施有效的法律规制。学界普遍认为，现行的以人为中心的国际人道主义法律框架和正义战争伦理原则，不能适用于智能机器人之间的战争。因此，将智能化武器装备视为道德行为者，需要改变现有的道德思维方式，构建以智能化机器人的道德思维方式为框架的伦理和法理体系。

就实现战争责任正义而言，只要战争不消亡，人类的痛苦还存在，维持人类对杀戮行为的责任追究在道义上就是合理和必要的。因此，必须在遵循责任匹配原则的基础上构建法律体系进行规制。

首先，明晰主体责任。针对智能化武器装备作战应用可能出现的“问责空白”和责任分配困境，应当在智能化武器装备的设计、研发、应用和发展战略制定等全过程中遵循责任主体匹配原则，厘清不同责任方在各环节中所应当承担的责任。政治家对国家智能化战争决策和军事智能化发展的战略性失误承担责任；军工企业和研发人员对反恐无人机设计制造失误及其造成的战争后果负责；军方参战人员应分别对战役、战术指挥和具体操作过程中违背战争规约的行为负责。

其次，制定和完善法规制度。在就各主体责任达成共识的基础上，国际社会应当借鉴两次海牙和平会议、《日内瓦公约》及其附加议定书的成功经验，形成一整套规范智能化武器装备使用部署的国

① SULLINS J P. When is a robot a moral agent? [J]. International Review of Information Ethics, 2006 (6): 23 - 30.

际法律框架和无人战争交战规约，作为约束智能化武器装备的设计、试验、制造、扩散和部署使用的法律依据。

最后，构建问责机制。国际社会需要共同磋商并建立一套行之有效的智能化武器装备使用问责机制，对任何未按照国际标准程序和惯例使用智能化武器装备从事军事行动，进而导致战争违规行为的组织和个人进行问责和追责。

四、国际军控规制进路

当前，美国在全球实施的无人机“反恐战争”愈演愈烈，智能化武器装备不仅是反恐作战的利器，而且成为美国在全球范围内推行强权政治的重要工具。因此，要对智能化武器装备引发的一系列社会问题进行有效规制，寄希望美国政府自身的努力可能是难以奏效的，必须要整个国际社会共同参与。当前美国滥用智能化武器装备进行反恐作战，武装无人机已经呈现出全球化扩散的趋势，由此也引发了国际社会对智能化武器装备进行军备控制的呼声。奥特曼等人指出：“最好的解决方案是利用当前的机会窗口，清晰地划设一条明亮的红线，反对自主武器系统，防止这些系统从人类手中夺走生死攸关的决定权。”① 考虑到智能化武器装备在作战中的滥用对人类尊严、生命权利、区域安全和世界长久和平可能带来的一系列不可预测的后果和难以承受的代价，国际社会必须共同努力，督促军事强国，并共同推动智能化武器装备在全球军备控制的进程。

智能化武器装备是人工智能和大数据技术应用于军事领域最前

① ALTMANN J, SAUER F. Autonomous weapon systems and strategic stability [J]. Survival, 2017, 59 (5): 134.

沿、最典型的体现，对其功能、作用及潜在风险的认识，存在较大分歧，不同行为主体基于自身利益诉求、认知观念、战略谋划和权力地位的差异，对智能化武器装备军备控制持有不同立场、意见和观点。为此，必须将智能化武器装备的军备控制贯彻到无人作战控制系统研制、开发和使用的全过程，依托联合国全面统筹协调各方诉求，切实加强国际合作，适时地、恰当地推进构建智能化武器装备的军控体系，通过友好协商的方式，就强化智能化武器装备的军备控制达成最广泛的战略共识，推动相关全球治理体系的搭建。为此，各国必须把握当前的战略机遇期，在联合国框架下共同推动军备控制的各项工作，共同敦促以美国为首的北约国家控制智能化武器装备的发展，具体体现在以下三个方面。

一是坚持联合国在智能化武器装备军备控制中的核心地位。智能化武器装备的军备控制，涉及研制、开发、扩散、部署和使用等方面，其牵涉国家、利益主体范围的广度，以及对全球未来安全格局影响的深度，使得以“促进国际合作和实现世界和平和安全”为宗旨的联合国具有其他国际行为主体无可比拟的优势，因此，必须始终坚持联合国作为全球性国际安全机构在维护智能化武器装备使用安全与全球战略稳定性中的核心地位，使之成为推动智能化武器装备军备控制规约制定和实施的中坚力量。

二是发挥联合国在智能化武器装备军备控制中的主导作用。借鉴国际社会禁止化学武器、生物武器等大规模杀伤性武器的有益做法，由联合国下属多部门和机构按照职责分工，全方位、深层次介入、推动国际合作。首先，依托联合国大会下属的联合国人权理事会这一保护人权的工作平台，就智能化武器装备作战应用对于人道主义原则和国际人权法的冲击与挑战达成广泛共识，形成启动作战规约磋商的认知基础。其次，推动联合国《特定常规武器公约》会

议成为智能化武器装备唯一的多边军控谈判机构，并将智能化武器装备的军备控制列入专题单元进行探讨，尽快推动智能化武器装备军控谈判由政府专家组会议逐渐上升到政府间谈判会议。依托联合国汇集国际社会力量，共同阻止当前智能化武器装备军备竞赛通向完全自主阶段的潜在趋势，最终签署具有法律约束力的预防性多边军备控制协议，全面禁止部署和使用具有自主武器系统的智能化武器装备。最后，借鉴联合国国际原子能机构的职能设置和工作机制，支持联合国设立智能化武器装备的专项核查机构，并赋予其具体的监督权力和核查使命，支持联合国建立维护智能化武器装备审查的国际透明机制，对各国最新研发的智能化武器装备、作战手段以及所涉及的材料设备进行严格而全面的审查，并评估其对于战场毁伤、法理、伦理乃至全球战略稳定性的潜在威胁，确保智能化武器装备研发和部署的合法性和可控性。

三是促进联合国在智能化武器装备的危机管控和升级管控中发挥更大作用。为应对智能化武器装备破坏交战规约进而带来国际安全危机快速升级的潜在风险，各国应当充分发挥联合国安全理事会维护国际和平与安全的职能定位。针对滥用智能化武器装备作战可能引发人道主义危机的风险，积极推动在联合国安理会设置安全分析小组，搜集各国智能化武器装备研发、部署和使用方面的最新信息，进行有针对性的调查研判，定期搜集智能化武器装备违反人道主义原则的行为，并提供风险评估报告。同时，考虑到智能化武器装备违规行为可能导致国际冲突中决策加速的特性，应当在联合国安理会框架下建立无人作战危机管控的应急机制，制定计划以应对突发情况。一旦智能化武器装备作战应用过程存在违规和失范行为，甚至引发国际人道主义危机的爆发，联合国秘书长和安理会能够在第一时间进行斡旋，协调各方立场，派遣军事观察员或派驻维和部

队，帮助缓解紧张程度，为和平解决争端提供方法或途径，尽可能地避免危机加速升级和冲突扩大化的风险。

此外，军控规制路径还包括推动大国之间约束智能化武器装备发展的战略合作。在以联合国为主导推进军备控制的基础上，拥有智能化武器装备技术的各个国家应当进一步深化交流合作，通过官方与非官方的对话、协商、谈判和热线交流，就智能化武器装备可能导致的战场失范行为和人道主义风险达成广泛共识，建立战略沟通与危机管控机制，推动达成实现维护智能化战争作战交战规约的协议、条约和法律体系，具体体现在以下三个方面。

一是形成战略共识。冷战期间，美苏两个超级大国盲目追求核打击能力与核威慑能力，迟迟未就核武器的战略稳定性达成共识，从而将整个世界笼罩在核战争阴影之下。当前，面对智能化作战方式日新月异的发展，各国应该吸取历史教训，充分认清滥用智能化武器装备以及放任其高度智能化发展的不确定性前景及对国际人道主义安全的一系列潜在威胁，共同提出维系智能化武器装备军备控制的要素、条件和框架。各国应当从核军备控制的曲折发展中吸取教训，为避免冷战时期“相互确保摧毁”的恐怖平衡再度出现，应共同致力于增加智能化武器装备研发的透明度，并将其置于严密的国际监管之下。同时，各国要充分认清恐怖分子、黑客组织掌控智能化武器装备技术对于全球安全的共同威胁，严控关键技术与关键设备向恐怖组织扩散的风险，共同构筑智能化武器装备扩散的全球监控网络，牢固确立使用智能化武器装备的底线意识，全面评估智能化武器装备部署和使用过程中的安全性与可靠性风险，确保将智能化武器装备技术的军事应用置于人类可以控制的范围之内，杜绝人道主义危机爆发和加速升级的潜在风险。

二是强化战略沟通。历史的经验与教训启示我们，封闭对峙无

法缓和国际关系，唯有交流对话才能促进了解、增强战略互信。当前，国际社会对于智能化武器装备潜在作战能力和使用意图存在巨大认知差异性和不透明性，需要各国进一步强化军备控制的战略沟通机制。首先，要强化美国等智能化武器装备技术先发国家之间的战略沟通，推动构建维护智能化武器装备军控机制，使之成为国际性或区域性安全峰会的重要议题，协调和管控各国智能化武器装备研发和部署行动，减少因信息沟通不畅而导致的战略误判，遏制不必要的危机升级风险。其次，要采取灵活的沟通方式，围绕智能化武器装备滥用行为引发的人道主义危机升级、地区恐怖主义活动加剧等一系列问题，采取滚动磋商形式和定期交流机制，努力在大国间构建由官方、非官方双向并行的交流机制，广泛吸纳战争法专家、武器技术专家、军备控制学者、军工企业家、战略研究专家、退役军官与退休外交人员参与1.5轨、2轨等非官方沟通机制，从技术、战略、经济、外交等多维视角广泛探讨维护智能化武器装备作战的交战规约所涉及的热点、难点问题，配合并服务于官方1轨磋商。最后，鉴于未来智能化武器装备作战应用的战场决策速度优势可能加剧人道主义危机快速升级的风险，可以考虑在将来可能出现的危机事件中，将智能化武器装备军备控制作为大国间热线交流的重要内容，以热线交流的直接性、快捷性来确保及时准确的信息沟通，从而降低因战略误判而导致的战争爆发和进一步升级的风险。

三是达成战略协议。考虑到短期内难以就智能化武器装备军备控制签署具有法律规范约束作用的国际法条文，拥有智能化武器装备技术的国家，特别是技术先发国家，应当在联合国《特定常规武器公约》官方专家组会议所达成共识的基础上，按照先易后难、逐步深入的总原则，围绕实现智能化武器装备逐步军备控制的目标，形成一系列具有可执行性的软法文书，通过行为准则、共同宣言、

框架公约和情报共享机制等多样化形式，逐渐推动达成双边、多边战略协定与合作内容，就“不使用智能化武器装备攻击核设施”“不使用智能化武器装备攻击太空资产”“不使用智能化武器装备攻击国家关键基础设施”“大规模杀伤性智能化武器装备试验通报机制”等具体议题，达成具有建设性的战略协议。

五、舆论规制进路

近年来，少数国家尝试以“零风险”“零伤亡”“低成本”等词语证明无人机“定点清除”行动的正义性，对此，国际社会必须保持清醒的认识，针对少数国家政府罔顾事实地试图为无人机反恐作战辩护的拙劣行径，相关主权国家、国际组织、民间社会力量应同心同德，充分利用国际舆论的力量，构建全球舆论规制网络，阐释无人机反恐作战引发的伦理风险和安全风险，揭露无人机反恐作战的非人道主义本质。

首先，在科学共同体内部营造遵循科研伦理和履行研发者主体责任的舆论氛围。为回应智能化战争可能对人类社会和平发展带来的威胁，必须呼吁和敦促科学家群体行动起来，从善良意志出发，进一步履行维护和平和关爱生命的伦理责任，具体体现为：做好前瞻性风险评估，敏锐洞察智能化技术军事应用可能带来的伦理风险与挑战，严格限制智能化作战系统的杀伤能力，通过道德代码设置和伦理嵌入等方式，按人类伦理、法律来规范智能化作战平台作战行为；坚决抵制滥用无人机反恐、不受限制的大数据反恐，禁止致命性自主武器系统的研发和应用，紧守人类掌控开火权的安全设计底线，规避智能化战争扩大化并酿成大规模人道主义灾难的潜在可能。

其次，发动整个国际社会形成对使用致命性自主武器系统的舆论攻势。在当前联合国框架下，针对维护智能化武器装备军备控制的官方谈判进程进展缓慢的问题，可以考虑利用“国际人工智能联合会议”“杀手机器人禁令运动”“未来生活研究所”等具有较大影响力的非政府组织力量，强化对智能化武器装备无序发展的国际舆论规制作用。上述非政府组织在智能化武器装备的舆论规制中可以发挥的作用具体体现为：一是通过电视、报纸、社交网络等媒体手段，公开揭露无人机反恐造成的人道主义危机，发起禁止研制和部署致命性自主武器系统的全民倡议，赢取民众对于无人作战违规行为全球治理的广泛支持；二是密切关注智能化武器装备研发前沿进展，利用专业知识优势提供战略咨询服务，就其潜在的人道主义风险发出预警信息并提供对策咨询；三是通过国际主流媒体舆论发声等形式，敦促以美国为首的相关国家和部门加快构建智能化武器装备作战违规和失范行为风险防控和追责机制，呼吁规范和限制智能化武器装备作战应用的国际立法。

最后，完善致命性自主武器系统的全球舆论规制行动者网络。智能化战争作战违规和失范行为的全球治理，需要汇聚国际组织、政府官员、科技专家、工程师、制造商、法律专家、军队指挥官、士兵和普通民众等众多行动主体的集体力量。构建全球舆论规制网络，必须在广泛交流沟通的基础上，为不同的行动主体设定明确的权利与责任。一方面，充分发动媒体向相关责任人宣传无人作战违规和失范行为可能引发的风险、危害；另一方面，利用舆论规制促进相关主体责任意识的觉醒，使其自觉限制智能化武器装备的滥用，自觉抵制智能化武器装备作战应用过程中的违规和失范行为。

需要指出的是，上述几条规制进路并不是孤立的，而是需要相互交叉和协力并进的。同时，以上讨论的技术、哲学、法律、军控

和舆论等规制进路也不能穷尽全部治理方案。唯有以未雨绸缪之心，唤起科学家、政治家、军事家、普通民众乃至整个国际社会的忧患意识，搭建起全球治理网络，才有希望在未来更从容地应对智能化武器装备带给人类社会生存、延续和发展的严峻挑战。

结　语

塞瑟等人指出："自主武器、人工智能、遥感、网络技术、高超音速飞行器、增材制造、隐身、精确制导和其他领域的重大进步使人们普遍认为，世界再次处于新技术的悬崖边。越来越多的人认为新兴技术有可能成为军事和战略事务中的'游戏规则改变者'。"[①] 面对全球人工智能军事化的发展浪潮，人类面临生存抑或毁灭的艰难抉择。如何实现人工智能领域的安全治理，有效应对风险挑战，成为各国军政领导人、人工智能专家、国际关系学者共同关注的议题。

一种观点认为，国际社会应当全面禁止以致命性自主武器为代

① SECHSER T S, NARANG N, TALMADG C. Emerging technologies and strategic stability in peacetime, crisis, and war [J]. Journal of Strategic Studies, 2019, 42 (6): 727.

和舆论等规制进路也不能穷尽全部治理方案。唯有以未雨绸缪之心，唤起科学家、政治家、军事家、普通民众乃至整个国际社会的忧患意识，搭建起全球治理网络，才有希望在未来更从容地应对智能化武器装备带给人类社会生存、延续和发展的严峻挑战。

结　语

塞瑟等人指出："自主武器、人工智能、遥感、网络技术、高超音速飞行器、增材制造、隐身、精确制导和其他领域的重大进步使人们普遍认为，世界再次处于新技术的悬崖边。越来越多的人认为新兴技术有可能成为军事和战略事务中的'游戏规则改变者'。"[①] 面对全球人工智能军事化的发展浪潮，人类面临生存抑或毁灭的艰难抉择。如何实现人工智能领域的安全治理，有效应对风险挑战，成为各国军政领导人、人工智能专家、国际关系学者共同关注的议题。

一种观点认为，国际社会应当全面禁止以致命性自主武器为代

① SECHSER T S, NARANG N, TALMADG C. Emerging technologies and strategic stability in peacetime, crisis, and war [J]. Journal of Strategic Studies, 2019, 42 (6): 727.

表的人工智能军事化进程，从而阻止人工智能成为下一个全球化的、全面开放的军备竞赛领域。奥特曼等人指出："最好的解决方案是利用当前的机会窗口，清晰地划设一条明亮的红线，反对自主武器系统，防止这些系统从人类手中夺走生死攸关的决定权。"① 全球人工智能专家共同签署《致命性自主武器宣言》，呼吁各国政府和领导人"创造强大的国际规范、法规和法律，以遏止未来致命性自主武器系统的发展，既不参与也不支持致命性自主武器的开发、制造、贸易或使用"。②

人工智能军事化进程方兴未艾，国际社会就全面彻底禁止研发、部署和使用致命性自主武器系统乃至全部智能化武器装备达成共识，共同签署一项多边的、具有法律约束力的预防性军控协议，不仅是政治上最为审慎、道德上最为正义的选择方案，而且可以一劳永逸地解决所有面临的或潜在的安全风险。然而，回顾人类军备控制的历史，我们应当清醒地意识到，只有当一种武器的现实危害远远超出它所展现的作战效用时，针对这种武器的禁令才可能真正生效。类似于核武器，人工智能可以在军事领域提供强有力的"不对称"战略优势。人工智能在军事领域部署和使用已经是不争的事实，因此所谓的红线已经被跨越。在当前人工智能的战争危害尚未完全显现的情况下，要求各军事强国放弃追逐具有广阔军事应用价值的人工智能武器，无疑是不现实的。正如沙瑞尔指出的："在需要全体协商达成一致意见的《特定常规武器公约》会议上，不可能通过针对自主武器的禁令。即便通过了，战时这些规则究竟能发挥多大效用，

① ALTMANN J, SAUER F. Autonomous weapon systems and strategic stability [J]. Survival, 2017, 59 (5): 134.

② Future of Life. Lethal autonomous weapons pledge [EB/OL]. (2018-06-06) [2022-04-07]. https://futureoflife.org/open-letter/lethal-autonomous-weapons-pledge/.

也值得商榷。”①

另一种观点则认为，人工智能领域的军备竞赛是不可避免的，基于道德、法律和战略安全等方面的考量，保持人类对人工智能军事化发展进程的控制权，就成为一个理性和现实的选择。盖斯特指出：“由于人工智能研究与军事应用之间的长期联系，放弃人工智能军备竞赛为时已晚。我们需要在管理有序、安全可靠的人工智能军备竞赛与管理不善、可能导致灾难性后果的军备竞赛之间做出选择。”② 当前，围绕人工智能如何实现可控、安全的军事化发展，存在多种治理方案。联合国《特定常规武器公约》会议致命性自主武器政府专家组提出，所有武器都应受到“有意义的人类控制”。③ 沙瑞尔提出，我们需要“创建保持人类判断在战争中发挥作用的通用准则”。④ 美国国防创新委员会发布的《人工智能原则：国防部人工智能应用伦理的若干建议》，就美军如何合乎伦理和法律地使用人工智能技术，提出“责任、公平、可追踪、可靠、可控”五大原则。⑤

上述方案在允许研发、部署人工智能武器的前提下，通过技术、政策等手段为人工智能武器确立行为准则，限制人道主义灾难的发生，其作为军备控制原则具有一定的宏观指导意义，但在实践层面

---

① 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：397.

② GEIST E M. It's already too late to stop the AI arms race—We must manage it instead [J]. Bulletin of the Atomic Scientists, 2016, 72 (5): 318.

③ RIEBE T, et al. Meaningful human control of Lethal Autonomous Weapon Systems: the CCW - Debate and its implications for VSD [J]. IEEE Technology and Society Magazine, 2020, 39 (4): 36.

④ 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：400.

⑤ Defense Innovation Board. AI principles: recommendations on the ethical use of Artificial Intelligence by the Department of Defense: supporting document [Z]. United States Department of Defense, 2019.

的可操作性是值得怀疑的。具体而言，上述方案的逻辑基点，就是要维持“人在回路”的军事系统结构，从而实现对人工智能系统的“有意义的人类控制”。然而，综合考虑人工智能系统和战争系统的双重复杂性和不可控性，在战争行为中，人工智能武器能否准确理解、严格执行人类指令本身就具有不确定性，人类也往往难以清晰地认识高度复杂的人工智能“技术黑箱”。在笔者看来，以自主武器系统为代表的智能化武器装备应用于战争，标志着人与武器关系层面的一次结构性转变，这种转变类似于马克思所指出的“机器对人的反向统治”。机器自治系统的创造意味着战争权力的转移，其中机器不再是人类用于特定目的的工具，相反，机器有自己的目的，人类的角色是监督和防范。在这种关系中，人的行为是由机器运动决定和调节的，因而只能亦步亦趋地追随机器。需要警惕的是，如果各国政府、军控专家对于类似“有意义的人类控制”的解决方案抱有高度自信，甚至将其作为军备控制协议或治理工作的基石，就可能会将人工智能军控引上危险的歧途。

当前联合国《特定常规武器公约》框架下的人工智能军控，从界定概念和提出治理原则出发，致力于签署全体缔约国家协商一致的军控协议，难以在短期内取得突破性进展。自 2017 年成立致命性自主武器政府专家组以来，迄今仍停留在相关概念原则的争论阶段，始终未被授权商讨多边军控协议。[①] 比较而言，相对于全体缔约国家共同参与的军控会谈和烦琐的概念辩论，先由少数技术先发国家围绕人工智能军事化的共同安全关切达成战略共识，再向其他缔约国家扩展，应当成为更具可行性的军控路径。一方面，虽然人工智能

① 沙瑞尔. 无人军队：自主武器与未来战争 [M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019：389.

技术具有军民一体化和易于扩散的特征，但是前沿领域人工智能创新、高端人力资源，以及强大的数据库和运算能力，仍然集中掌握在少数技术先发国家手中，防止尖端人工智能技术及其军事应用向技术后发国家扩散，符合技术先发国家的安全利益诉求，易于达成少数国家之间的防扩散技术协议。另一方面，技术先发国家在人工智能军备发展中可能面临与“核禁忌”同等强度或更为强烈的“人工智能禁忌”。例如，隐形无人机对国家核设施实施“蜂群”攻击，可能打破现有的攻防战略平衡，引发有核国家的恐慌情绪；人工智能与核指挥控制系统的结合，可能造成核战争意外爆发和无限升级的风险；人工智能作为网络攻击手段，可能导致全球网络空间和能源、通信等关键基础设施的瘫痪。围绕上述共同安全关切达成战略互信，符合技术先发国家的共同安全利益诉求。

同时，鉴于由主权国家主导的自上而下的模式在推动人工智能军控方面进展缓慢，还可以考虑自下而上的军控进路，即由知识精英形成对某种军备发展危害性的“认知共同体”，进而塑造规范，并以自下而上的知识传播方式影响、促进主要国家间达成军备控制协议。回顾核军备控制的历史，1972 年美苏签署《反弹道导弹条约》为我们提供了一个知识精英自下而上塑造规范的军控范例。[①] 当前在人工智能领域，知识精英已经初步具备了构建军控“认知共同体”的现实条件。首先，马斯克、扎克伯格、哈萨比斯等业内顶尖专家已经就人工智能军事化发展的潜在危害积极发声。如马斯克表示：“人工智能可能是引发第三次世界大战的最有可能的原因。”“人工

① ADLER E. The emergence of cooperation: national epistemic communities and the international evolution of the idea of nuclear arms control [J]. International Organization, 1992, 46 (1): 101.

智能是人类文明存在的根本风险。”[①] 其次，全球范围内的人工智能专家就人工智能军事化发展的风险达成了共识，形成了“杀手机器人禁令运动”“未来生活研究所”“国际人工智能联合会议”等国际性互动社区，并以签署《致命性自主武器宣言》等形式向全球传播。目前，包括中国、巴西、阿根廷在内的30多个国家明确支持禁止致命性自主武器系统。[②] 最后，人工智能领域的前沿技术和创新资源集中于知识精英密集的私营部门，基于文化传统、价值观念和风险认知，知识精英可以在与政府、军方开展的合作中发挥重要作用，甚至影响人工智能研发进程和政策规范的形成。例如，2018年5月，3100名谷歌员工以“不作恶”为由签署抗议信，迫使公司退出美国国防部无人机自主识别项目。[③] 基于类似的理由，谷歌DeepMind、波士顿动力和沙夫特等人工智能公司也取消了与美国政府签订的项目合同。[④] 上述事件直接推动了美国国防部出台《人工智能原则：国防部人工智能应用伦理的若干建议》。

中国作为联合国安理会五大常任理事国之一，在继续参与联合国框架下军备控制谈判的同时，也可以发挥大国影响力，积极投身自下而上的人工智能安全治理进程，汇聚全球热爱和平的科学家、非政府组织和民间团体的合力，共同构建全球治理网络。针对人工

① Elon Musk Neuralink. Elon Musk's concerns on Artificial Intelligence (AI) [EB/OL]. (2021-10-12) [2022-04-07]. https://elonmuskneuralink.com/elon-musks-concerns-on-artificial-intelligence-ai/.

② Future of Life. Lethal autonomous weapons pledge [EB/OL]. (2018-06-06) [2022-04-07]. https://futureoflife.org/open-letter/lethal-autonomous-weapons-pledge/.

③ HOLLISTER S. Project Maven: nearly a dozen Google employees have reportedly quit in protest [EB/OL]. (2018-05-14) [2022-04-07]. https://www.cnet.com/tech/tech-industry/google-project-maven-drone-protect-resign/.

④ Congressional Research Service. Artificial Intelligence and National Security [R]. Congress, 2019: 19.

智能军事应用产生的安全、伦理和法律问题，可以组织科技专家、战略专家、伦理学家和法学家等知识精英通过主流媒体、门户网站和社交网络等渠道积极发声，面向全球发布人工智能军事应用的科研伦理倡议，开展关于人工智能军事应用风险的科学、伦理和法律知识普及，并就约束研发、部署人工智能武器行为发布官方立场文件，在全球范围内塑造、构建并不断强化人工智能军事应用风险的共有认知，赢得民众对人工智能军事化发展进行全球治理的广泛支持，以强有力的国际舆论压力敦促各国政府加快推进人工智能多边军控协议的进程。

参考文献

一、专著类

[1] 耶讷. 无人机攻击[M]. 刘春辉，赵希源，译. 北京：中国市场出版社，2014.

[2] 马泽蒂. 美利坚刀锋：首度揭开无人机与世界尽头的战争[M]. 王祖宁，王凌凌，美同，译. 北京：新世界出版社，2014.

[3] 沙瑞尔. 无人军队：自主武器与未来战争[M]. 朱启超，王姝，龙坤，译. 北京：世界知识出版社，2019.

[4] 沃尔泽. 正义与非正义战争：通过历史实例的道德论证[M]. 任辉献，译. 南京：江苏人民出版社，2008.

[5] 布特. 战争改变历史：1500年以来的军事技术、战争及历史进程[M]. 石祥，译. 上海：上海科学技术文献出版社，2011.

[6] 卡洛，弗鲁姆金，克尔. 人工智能与法律的对话[M]. 陈吉栋，董惠敏，

杭颖颖，译. 上海：上海人民出版社，2018.

[7] 考彼尔特斯，福臣，时殷弘. 战争的道德制约：冷战后局部战争的哲学思考[M]. 邹琳，戴锋宁，等译. 北京：法律出版社，2003.

[8] 王铁崖. 国际法[M]. 北京：法律出版社，1981.

[9] 周桂银. 国际政治中的外交、战争与伦理[M]. 南京：南京大学出版社，2018.

[10] 左高山. 战争镜像与伦理话语[M]. 长沙：湖南大学出版社，2008.

[11] 乔良，王湘穗. 超限战[M]. 北京：中国社会出版社，2005.

[12] 朱之江. 现代战争伦理研究[M]. 北京：国防大学出版社，2002.

[13] KRISHNAN A. Killer robots: legality and ethicality of autonomous weapons[M]. London: Routledge, 2009.

[14] BHUTA N, BECK S, GEISS R, et al. Autonomous weapons systems: law, ethics, policy[M]. Cambridge: Cambridge University Press, 2016.

[15] SINGER P W. Wired for war: the robotics revolution and conflict in the 21st century[M]. London: Penguin Books, 2009.

[16] AKHGAR B, SAATHOFF G B, ARABNIA H R. et al. Application of big data for national security[M]. Oxford: Butterworth-Heinemann, 2015.

二、论文类

[1] 傅莹. 人工智能对国际关系的影响初析[J]. 国际政治科学，2019，4(1)：1－18.

[2] 封帅. 人工智能时代的国际关系：走向变革且不平等的世界[J]. 外交评论，2018，35(1)：128－156.

[3] 刘晓娟，李志，王彤. 2019年美国空军人工智能附录：响应国防部人工智能战略[J]. 无人系统技术，2019，2(4)：59－61.

[4] 阙天舒，张纪腾. 美国人工智能战略新动向及其全球影响[J]. 外交评论，2020，37(3)：136.

[5] 周琪，付随鑫. 美国人工智能的发展及政府发展战略[J]. 世界经济与政治，2020(06)：29.

[6] 盖伊斯特，罗恩. 人工智能会怎样影响核战争风险？[J]. 国外社会科学文摘，2018(8)：16.

[7] 张煌，杜雁芸. 人工智能军事化发展态势及其安全影响[J]. 外交评论，2022(3)：116.

[8] 张煌，杜雁芸. 全球自主武器系统的发展及其对战略稳定性的影响[J]. 国际论坛，2021，23(3)：79.

[9] 徐能武，葛鸿昌. 致命性自主武器系统及军控思考[J]. 现代国际关系，2018(7)：56.

[10] 钱铖，石斌. "雄蜂"的阴影：美国无人机作战对当代战争伦理的挑战[J]. 世界经济与政治，2013(8)：86－99.

[11] 黄云松，蔡瑞艳. 无人攻击机面临的国际法挑战：以美国在巴基斯坦的无人机攻击为例[J]. 南亚研究季刊，2012(1)：21－25.

[12] 刘正. 网络战的国际法应对[J]. 山东社会科学，2006(3)：139－141.

[13] 张煌，朱启超，李坡. 信息化战争视阈下武器演进的伦理悖论[J]. 中国军事科学，2013(5).

[14] 何奇松. 美国网络威慑理论之争[J]. 国际政治研究，2013(2)：58－59.

[15] 王飞跃. 面向赛博空间的战争组织与行动：关于平行军事体系的讨论[J]. 军事运筹与系统工程，2012(9)：7.

[16] 梁亚滨. 武装无人机的应用：挑战与影响[J]. 外交评论，2014(1)：143－156.

[17] 张煌，刘轶丹，张鹏. 无人机作战的伦理困境[J]. 伦理学研究，2015(2)：133－137.

[18] KOLFF D W. Missile strike carried out with Yemeni Cooperation—using UCAVs to kill alleged terrorists: a professional approach to the normative bases of military ethics[J]. Journal of Military Ethics，2003(3)：240－244.

[19] McCormick, Patrick. Attack of the drones[J]. U. S. Catholic，2005(8)：46

-48.

[20] LEKEA, JOANNE K. Missile strike carried out with Yemeni cooperation—the war against terrorism: a different kind of war? [J]. Journal of Military Ethics, 2003(3):230-239.

[21] AHMAD M. The use of drones in Pakistan: an inquiry into the ethical and legal issues[J]. Political Quarterly, 2014(1):65-74.

[22] BOYLE M J. The legal and ethical implications of drone warfare[J]. The International Journal of Human Rights, 2015(2):105-126.

[23] SCHWARZ E. Prescription drones: on the techno-biopolitical regimes of contemporary ethical killing [J]. Security Dialogue, 2016(1):59-75.

[24] VORSTER N. Killing from a distance: a christian ethical evaluation of CIA targeted drone killings[J]. The Heythrop Journal, 2015(5):836-849.

[25] BYRNE E. Making drones to kill civilians: is it ethical? [J]. Journal of Business Ethics, 2018(1):81-93.

[26] ENEMARK C. Drones over Pakistan: secrecy, ethics, and counterinsurgency [J]. Asian Security, 2011, 7(3):218-237.

[27] KNOOPS G J A. Legal, political and ethical dimensions of drone warfare under international law: a preliminary survey[J]. International Criminal Law Review, 2012, 12(4):697-720.

[28] KARPPI T, BOHLEN M, GRANATA Y. Killer robots as cultural techniques [J]. International Journal of Cultural Studies, 2018(3):107-123.

[29] BOOGAARD V D, JEROEN. Proportionality and autonomous weapons systems [J]. Journal of International Humanitarian Legal Studies, 2015, 6(2):247-283.

[30] DIPERT R R. The ethics of cyberwarfare[J]. Journal of Military Ethics, 2010, 9(4):384-410.

[31] LIN P, ALLHOFF F, ROWE N C. War 2.0 cyberweapons and ethics[J]. Communications of the Acm, 2012, 55(3):24-26.

[32] COOK C J. Cyberation and just war doctrine:a response to randall dipert[J]. Journal of Military Ethics,2010,9(4):411 -423.

[33] DIPERT R R. Other-than-Internet cyberwarfare: challenges for ethics, law, and policy[J]. Journal of Military Ethics,2013,12(1):34 -53.

[34] DANKS D, DANKS J H. The moral permissibility of automated responses during cyberwarfare. [J]Journal of Military Ethics, 2013,12(1):18 -33.

[35] OREND B. Michael walzer on resorting to force Canadian [J]. Canadian Journal of Political Science,2000,33(3):523 -547.

[36] DIPERT R R. Preventive war and the epistemological dimension of the morality of war [J]. Journal of Military Ethics,2006,5(1):32 -54.

[37] DIPERT R R. The ethics of cyberwarfare [J]. Journal of Military Ethics, 2006,9(4):84 -410.

[38] ROWE N C. The ethics of cyberweapons in warfare[J]. International Journal of Technoethics,2010,1(1):20 -31.

[39] GALBRAITH J . U. S. Drone strike in Iraq kills Iranian military leader Qasem Soleimani[J]. American Journal of International Law, 2020, 114(2): 313 -323.

[40] LEVERINGHAUS A. What's so bad about killer robots? [J]. Journal of Applied Philosophy,2018,35(2):341 -358.

[41] LWEIS G N,POSTOL T A. How U. S. strategic antimissile defense could be made to work[J]. Bulletin of the Atomic Scientists,2010,66(6): 8 -24.

[42] GOOSE S. The case for banning killer robots: point[J]. Communications of the ACM,2015,58(12):43 -45.

[43] ALTMANN J, SAUER F. Autonomous weapon systems and strategic stability [J]. Survival,2017,59(5):117 -142.

[44] SCHULZKE M. Autonomous weapons and distributed responsibility[J]. Philosophy & Technology,2013,26(2):203 -219.

[45] SULLINS J P. When is a robot a moral agent? [J]. International Review of

Information Ethics,2006,12(6):23 - 30.

三、报告类

[1] GUETLEIN M A. Lethal autonomous weapons—ethical and doctrinal implications[R]. Naval War College,2005.

[2] DINAUER S R,SWEENEY P. The joint force commander's ethical challenges for the 21st century[R]. Naval War College,2006.

[3] DAVI L E,MCNERNEY M J, CHOW J,et al. Armed and dangerous? UAVs and U.S. security[R]. RAND Corporation,2014.

[4] DAHM W J A. Report on technology horizons:avision for Air Force science and technology during 2010 - 2030 [R] . Office of the Chief Scientist of the U. S. Air Force. 2010.

[5] ALSTON P. Report of the special rapporteur on extrajudicia, summary or arbitrary executions[R]. United Nations General Assembly,2010.

附　录

作者发表的本书主题相关论文

[1] 张煌，杜雁芸. 俄美军用人工智能竞争的战略稳定风险及其治理路径[J]. 俄罗斯研究，2022(6)：157－190.

[2] 张煌，林涵，张恺悦. 智能化战争的伦理风险与治理路径[J]. 中国军事科学，2022(5)：109－117.

[3] 张煌，杜雁芸. 人工智能军事化发展态势及其安全影响[J]. 外交评论，2022，39(3)：99－130.

[4] 张煌，杜雁芸. 人工智能时代的“数据霸权”与战争正义性问题审辨[J]. 思想理论战线，2022，1(6)：92－98＋141.

[5] 张煌，杜雁芸. 自主武器系统对当代战争伦理的挑战[J]. 社会科学，2021(4)：35－47.

[6] 张煌，贾珍珍. 无人作战的人机融合：挑战与出路[J]. 国防科技，2020，

41(6): 105 - 109.

[7] 张煌, 黄关. 大数据时代的分配正义挑战与规制原则[J]. 山东科技大学学报(社会科学版), 2020, 22(2): 22 - 27 + 54.

[8] 张煌. 中国海军识别查证美军无人潜航器的正义性分析[J]. 太平洋学报, 2019, 27(2): 83 - 92.

[9] 张煌, 刘轶丹. 网络战争伦理: 回顾与展望[J]. 伦理学研究, 2018(2): 124 - 128 + 134.

[10] 张煌, 冷妹, 古丽娜. 美军无人机作战的三大伦理困境[J]. 装备学院学报, 2016, 27(3): 22 - 27.

[11] 张煌, 朱启超. 论网络武器对现代战争伦理体系的挑战[J]. 自然辩证法研究, 2015, 31(11): 98 - 103.

[12] 张煌, 刘轶丹, 张鹏. 无人机作战的伦理困境[J]. 伦理学研究, 2015(2): 133 - 137.

[13] 张煌, 朱启超, 吴彤. 网络武器正义性的实践困境[J]. 装备学院学报, 2015, 26(1): 22 - 26.

后　记

《智能化战争伦理：风险与应对》一书即将出版，这是继《变化的规约：高新边疆的战争伦理》之后，我撰写的第二本关于战争伦理的专著，也是对我所承担的国家社会科学基金青年项目“美国无人机反恐的正义性研究”课题研究报告及近年来在《社会科学》《伦理学研究》《中国军事科学》等高水平期刊发表的一系列学术论文的系统梳理总结。

回首近年来沿着战争伦理方向一路前行的峥嵘岁月，不禁感慨万千！我研究生阶段所学专业并非伦理学，而是科学技术哲学。11年前，我于国防科学技术大学科学技术哲学专业博士毕业时，开始慎重思考今后工作阶段的科研方向：是继续从事科学技术哲学的研究，还是开拓一个新的研究领域和研究方向？这是当时我苦苦思索的一个问题。

科学技术哲学是一门百年学科。它的发展直接得益于19世纪末20世纪初的物理学革命。当时，面对经典物理学大厦上漂浮的三朵乌云，以爱因斯坦、玻尔、海森堡、狄拉克、玻恩、德布罗意等人为代表的一批兼具高深的理论造诣和深邃的哲学思维的物理学家，他们相互协作，共同推动了理论物理学从宏观的、低速的牛顿力学体系向微观的、高速的新物理学体系发生“范式”转变。作为哲学领域对科学革命的回应和反思，20世纪上半叶，科学哲学领域涌现了以库恩、拉卡托斯、波普尔、石里克、维特根斯坦、费耶阿本德、劳丹等人为代表的一批杰出的科学哲学专家，他们所著的《科学革命的结构》《猜想与反驳》《科学革命纲领研究方法论》《必要的张力》等著作，成为轰动20世纪哲学史的学术经典，至今仍然是代表这一学科领域最高水平的智慧结晶。然而，科学哲学经历了几十年的辉煌之后，就开始进入发展的低谷期，缺乏新生的学术大师和足够分量的研究成果。究其原因，主要是由于科学技术哲学学科的发展高度依赖自然科学的革命性突破，而自然科学的发展自19世纪末、20世纪初科学革命之后，在整个20世纪没有孕育真正意义上代表“范式”转变的新的科学革命，缺乏与爱因斯坦、海森堡、玻尔等人相匹敌的科学大师，缺少与相对论、量子力学相媲美的科研成果，自然也就无法带动科学技术哲学领域的发展创新。

相较于科学技术哲学100多年的漫长发展历程，战争伦理是一门20世纪70年代以来才真正诞生的新兴学科。虽然人类历史上几千年来一直不乏具有朴素战争伦理观念的思想家，但是真正战争伦理学科的建立，还是要回溯到20世纪70年代以后，以哈佛大学政治学教授沃尔泽为代表的一批伦理学者，系统地反思了美军在越南战争中存在的一系列非人道主义的行为，提出了以开战正义、交战正义和战后正义为核心的战争伦理研究基本框架，并引领了这一学

科的发展。

与科学技术哲学高度依赖于科技革命的发生相类似，战争伦理的研究也与人类的战争史紧密关联。冷战结束以来，伴随两极化的权力格局被打破，全球安全环境加剧动荡，以海湾战争为代表的局部战争频繁发生，由此也引发了中外学界关于高技术战争条件下战争伦理问题的新一轮思考。中国学者乔良、王湘穗在其代表作《超限战》中，旗帜鲜明地提出了信息化战争将会导致战争不断“慈化”这一极具争议性的论断，从而进一步推动了国内外学界对于军事技术与战争伦理互动作用的广泛论辩与深入思考。特别是在21世纪初，美国为谋取全球霸权，先后发动了阿富汗战争、伊拉克战争、利比亚战争等一系列高技术条件下局部战争，在全球范围内发起以打击“基地”组织等恐怖主义势力为名义的“反恐战争”，推动无人机、网络武器、太空武器等高新武器装备的作战应用，将这一领域的研究进一步引向深入，高水平的理论创新成果不断涌现。

我能够因缘际会地进入这一处于高速发展阶段的研究领域，得益于研究生阶段学科点开放交融的学术氛围和学科点老师对于战争伦理前沿的高度关注。学科奠基人刘戟锋教授在1998年创立国防科学技术大学科学技术哲学学科时，就把“科学品位至上、军事特色鲜明”作为学科发展的基本定位，大胆突破传统科学技术哲学研究范围的局限，从科学、技术、军事、哲学的交叉视野进行创新研究，刘戟锋教授早在1994年就于《伦理学研究》杂志上发表了《战争伦理：一种世界观念》的文章，该文章成为国内战争伦理研究的标志性成果。黄甫生教授和曾华锋教授合著的《武器的悖论》一书，系统归纳概括和总结了装备伦理领域的理论研究成果和前沿进展。在攻读博士学位阶段，这些老师、这些思想、这些创新成果对我产生了很大影响，为我走向战争伦理研究的道路埋下了种子。

2012 年 6 月，我留校工作，在国防科技大学国家安全与军事战略研究中心担任助理研究员，“立足国家安全、聚焦未来战争”的研究定位使我下定决心，离开耕耘多年、早已驾轻就熟的哲学象牙塔，以一个哲学专业博士生的独特视角研究高新技术战争中的伦理问题。

此后，岁月匆匆，我经历了初期筚路蓝缕的艰难开拓，也经历了游离于伦理学主流之外的挣扎彷徨，现在还需要应对学院调整后教学任务更加繁重的挑战。这十余年里，虽然我也曾经有过动摇和松懈的时候，但始终坚持研究的初心，以“任凭风浪雨打，我自岿然不动”的毅力，在学术研究的小径上踽踽独行。好在还有亲人的期待、师长的鼓励、学生的信任，让我不感无聊和孤独。

感谢我的父母，他们在并不富裕的条件下没有放弃我的学业，给我树立了读书改变命运的信念；感谢我的妻子和孩子对我的工作的支持，虽然他们并不清楚我写了什么，但是总感觉我在从事高尚的事业，其实这些年我在事业上碌碌无为，家庭方面也亏欠很多；感谢我的恩师朱亚宗教授，他总是鼓励我，在我失落和彷徨的时候不断开导我，希望有一天能够写出一本让他满意和骄傲的著作。感谢刘戟锋教授、曾华锋教授、石海明教授，他们引领我走上战争伦理研究的道路。感谢国防科技大学出版社的周蓉编辑和其他工作人员，正是他们的辛勤努力，本书才得以顺利出版。

愿未来的世界没有战争！

张 煌

2023 年 6 月于长沙