

致命性自主武器系统议题下有意义的人类控制原则

冷新宇

【摘要】《常规武器公约》框架下政府专家组会议对致命性自主武器的讨论,确定了有意义的人类控制原则的中枢地位。有意义的人类控制原则之所以被提出,是因为人工智能带来武器系统自主化的背景下,使用武器人的行为与武器系统攻击存在不一致性,后者并不一定是前者的必然结果。为消除伦理危机,实践中在两个方面阐述有意义的人类控制原则:一方面从军控角度出发,建立自开发至使用全流程监管,以期实现武器攻击反映操作员、指挥官意图的目标;另一方面从国际刑法、国家责任角度出发,落实人(国家)的责任不得被转移至机器的原则。有意义的人类控制原则在约束国家行为以及追究个人刑事责任、国家责任两个层次发挥作用,对我国法律的完善及未来塑造规制致命性自主武器的国际法规则,均有重要参考意义。

【关键词】有意义的人类控制原则;致命性自主武器系统;第36条审查

【作者简介】冷新宇,法学博士,中国政法大学法学院副教授。

【原文出处】《国际法学刊》(京),2022.2.1~36

【基金项目】本文是国家社科基金2019年年度项目“人工智能背景下自主武器系统相关国际法问题研究”(19BFX209)的研究成果。

人工智能技术在军事领域的运用是国际法和军控领域近年来的热点话题,致命性自主武器系统就在这一背景下提出。1980年《禁止或限制使用某些可被认为具有过分伤害力或滥杀滥伤作用的常规武器公约》(Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to be Excessively Injurious or to Have Indiscriminate Effects,以下中文简称为《常规武器公约》,英文简称为“CCW”)缔约国审议会议自2014年开始讨论致命性自主武器的概念和法律问题,2017年后缔约国同意每年召开政府专家组会议,讨论未来规制致命性自主武器^①的法律规则,2019年沙特阿拉伯阿美石油公司油田遭胡塞武装无人机袭击,2020年阿塞拜疆与格鲁吉亚争夺纳卡地区的武装冲突中大量使用土耳其产TB-2无人机,表明自主武器系统正在改变作战方式。

运用人工智能技术是否会使武器系统脱离人类

的控制,是军控和国际人道法、国际刑法领域的重要问题。2014年,红十字国际委员会提出对武器系统的关键功能施加有效的人类控制的理念,得到比较广泛的认同。^②美国国防部于2012年发布3000.09号指令,规范自主武器系统的发展、测试、部署和使用,保持指挥官或操作员对自主或半自主武器系统恰当程度的控制是其首要政策目标,^③前述指令使得组成自主武器系统硬件、软件有足够的鲁棒性(Robust),从而将非指挥官、操作员意图实施的交战或丧失对系统控制而导致的损失最小化。^④为落实美国国防部2018年人工智能战略,国防创新委员会(Defense Innovation Board)在其倡议文件《人工智能原则:关于国防部符合伦理的适用人工智能》(AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense)中提出了落实军事人工智能的五项原则:责任原则、公平原则、可追溯性原则、可靠性原则、可控制原则。^⑤此后,红十字

国际委员会在2021年向致命性自主武器议题政府专家组(Group of Governmental Experts, 以下简称“GGE”)会议提交的立场文件称:“自主武器系统的使用者不会对具体的目标、使用武力的准确时间或地点做出选择。这一过程的风险是丧失对武力使用的人为控制。这是人道、法律和道德关切的根源。”^⑥上述文件表达了对人工智能技术军事运用的伦理关切,其对应的军控规则和国际人道法要点在历次GGE会议中进入详细讨论。

2018年GGE会议以报告形式确认了致命性自主武器的开发、测试、部署及使用过程中应当贯彻的10项指导性原则,2019年GGE会议增加至11项,其中与有意义的人类控制原则有关的内容包括:(1)应确保根据适用的国际法在《常规武器公约》的框架内对发展、部署和使用任何新武器系统问责,包括使这类系统在人类指挥和控制的责任链中运作;(2)风险评估和减小风险的措施应成为任何武器系统新技术的设计、发展、测试和部署周期的组成部分;(3)可以采取各种形式,并在武器生命周期的各个阶段实施的人机交互,应确保对基于致命性自主武器系统领域的新兴技术武器系统的潜在使用符合适用的国际法,特别是国际人道法。在确定人机交互的质量和程度时,应考虑包括作战环境,以及整个武器系统的特点和能力等因素;(4)在发展或取得基于致命性自主武器系统领域新技术的新武器系统时,应采取适当的物理或非物理的防扩散风险控制措施;(5)人类对自主武器使用产生的后果负有责任,而非将人的责任转移至机器。^⑦上述内容约可统辖在有意义的人类控制原则(Principle of Meaningful Human Control)之中。

2021年GGE会议期间,各国针对指导性原则提交立场或评论文件,内容上大体支持上述共识。GGE将上述共识归纳为“规制自主武器”原则,各主要军事大国都倾向折中观点,甚至有国家认为合理使用自主武器将提升作战效能,减轻战争受难者的不必要痛苦,有助于交战国更好地遵守国际人道法。因此,尽管CCW框架下讨论的未来发展方向,可能走向军控条约为代表的硬约束,或者走向使用指南、法律编纂文本等为代表的软约束,但是,将自

主武器系统置于有意义的人类控制之下,是主要军事大国的基本共识。深入挖掘并富有意义的人类控制原则的意涵,是未来军控规则和国际人道法、国际刑法面临的重要问题。

本文首先分析为何应当坚持有意义的人类控制原则,随后讨论其在行为规则和责任规则中应如何体现,以期在自主武器规则的塑造中如何施加中国的影响力提供参考。

一、问题背景:操作员行为与武器攻击的不一致性

(一)操作员行为与武器攻击不一致性的技术原理

有意义的人类控制原则的产生背景,是自主武器的攻击不能一概视为操作员行为的延伸或必然后果,核心问题是武器系统在激活后基于其自主判断而“决策”发动攻击,与操作员、指挥官的行为在客观上是否具有因果关系,主观上能否反映操作员、指挥官的意图。

对于这一问题的框架,军事学上通常由观察、评估、决策、攻击四个可循环步骤构成的“回路”予以分析。前人工智能时代,回路中每一个步骤都由人来完成;人工智能技术引入后,每一个步骤均可由计算机系统完成,尤其是决策和攻击两个步骤交由计算机完成后,武器系统具有完全的自主性。武器系统自主决策后发动的攻击,是否被视为操作员的行为,由此产生了武器动作不一定等同于法律意义上行为的讨论。通过一系列“假设……然后……”的代码表达的复杂逻辑关系,能让武器系统具备初级自主判断能力;而在自主性更高的发展阶段,神经网络系统结合算法设计,可实现机器深度学习,即在输入端与输出端之间,添加数量可增加的虚拟层,而对机器的训练只需在输出端设定最优化导向的目标,通过“喂养”大量数据使计算结果不断趋近于设定目标。^⑧武器系统自主性地实现原理表明,系统的自主判断并不是人类智能思考的结果,仅是看起来比较像人类思考后做出决定的结果。

在国际人道法和国际刑法的实践中,从未出现过自主性引起的人的行为与被使用武器攻击后果不一致的情形。在先实践面对的是武器的误差问题。

克罗地亚军队在波黑战争中的暴风行动(Operation Storm)采用包括130毫米、155毫米榴弹炮、BM-21火箭炮来进攻军事目标,行动中同时轰击了城镇。审判庭根据专家证言判定,超出200米半径误差范围即为攻击违法。^⑨上诉庭拒绝接受审判庭的观点,理由是审判庭并未合理解释建立200米半径最大误差标准的基础,以及专家意见之间存在相互矛盾之处(有一位专家认为最大误差范围可以达到400米半径,审判庭没有解释为何选择最大允许误差范围为200米而非400米)。^⑩但上诉庭并没有否认审判庭考虑犯罪行为人与武器之间关系的前提性,即假定武器攻击动作是人的行为延伸,因而具有国际人道法、国际刑法原理中所界定人的“行为”的意义,犯罪行为人对需要对其触发武器的所有后果负责,武器攻击产生的误差也属于犯罪行为引发后果的范围。

自主武器可分为人在回路之中(man in the loop)、人在回路之外(man out of the loop)两种基本形态,二者对应不同的设计思路:前者尽管是自主性武器,但武器系统被激活后,操作员始终对回路保留控制权,可随时关闭或令武器系统自毁;后者通常被称为完全自主武器,激活后操作员对武器系统不再保留控制权。但上述分类仅仅是设计思想和武器类型的区分,实际情况是即便按照人在回路且对武器系统始终保持控制权的思路设计,自主性都将导致操作员行为与武器攻击出现不一致。原因可归纳为三点:

第一,高带宽、低延时、大容量的通讯链路无法建立或保证。来自地形、工作介质等作战环境因素的限制(如高频电磁波在水下无法长距离传播,而长波在水下可以传播但带宽极小无法加载更多数据),现有技术条件下无人潜航器若带有作战功能,很难在操作员与潜航器之间建立远距离、低时延的通讯链路。此外,作战环境下军事强国常常实施高强度电磁干扰,不能保证操作员时刻在回路之中控制。红十字国际委员会认为,保持人对回路的控制的三个基础性条件包括:态势感知能力、介入前的充足的反应时间、必要情况下夺回控制权或关闭武器系统的干预机制(通信连欧或者武力控制开关)。^⑪

第二,机器学习的算法黑箱问题。黑箱过程不能提供输出结果的合理性解释,对于战场态势感知

至关重要的计算机视觉和图像识别,机器学习的结果既不能被预见,也不能被解释。^⑫

第三,自主武器的设计和制造无法采用民用工业的技术迭代方法。民用产品可采集用户数据快速实现技术迭代,在较短时间内改进设计、消除不确定性。但武器及弹药造价昂贵,武器属于天然的“小数据”领域,智能化武器系统更是如此。

(二)法律规则复杂性对不一致性的放大

国际人道法对交战的规制存在一定的模糊性,表现在区分原则、比例原则以及预先警告原则的使用,此种模糊性放大了不一致性。

区分原则要求攻击对象限制在战斗员、军事目标的范围之内。理想条件下,图形识别技术、语音识别技术的进步有助于将平民和战斗员区分开来,从而提高攻击目标的准确性;但在区分原则的适用中,直接参与敌对行动的平民并不享有不受攻击的保护地位。^⑬技术上的难点是:一方面参与敌对行动的平民并不穿着制服,自主武器系统利用图形识别技术做出判断存在困难;另一方面,何为“直接参与敌对行动”(direct participation in hostility)具有复杂的法定构成要件,构成要件被注入了因果关系的要求,^⑭既有人工智能技术没有显示出此方面强于人类判断的能力。区分原则判断的另一困难是战斗员与平民混合在一起、军事目标与平民物体位置上极为靠近的状况。尽管图形识别技术有助于自主武器有效区分前述情况从而选择合适弹药精准攻击,但并非在任何情况下,武器系统都配备合适的弹药类型,武器攻击伤及无辜仍有可能(中性论者支持自主武器系统开发中加入攻击前警告的软硬件选择,以便提高自主武器的遵守区分原则可靠性)。^⑮基于区分原则的具体规则,在战场环境中的适用包含一系列复杂的逻辑和环境的判断。典型规则如日内瓦公约的《第一附加议定书》(以下简称为“《第一附加议定书》”)第56条原则上禁止对包含危险力量的物体、工程(如堤坝)进行攻击;但该条第2款又允许有例外情况,“如果该堤坝是用于其通常作用以外的目的和用以使军事行动得到经常、重要和直接支持的,而且如果这种攻击是终止这种支持的唯一可能的方法”,即允许进行攻击。上述规则包含复杂的战场环境评估、价值

判断要求,即便是战地指挥官也未必能在第一时间做出正确判断。在人工智能自主武器系统具备足够的观察感知能力、知识迁移能力、宏观比较能力之前,很难确定武器系统能达到人类判断水平。又如国际人道法中禁止无军事必要毁损财产的规则(1907年《海牙第四公约》附件《陆战章程》第23条g款),适用条件包含认定何为急迫的军事必要(imperative military necessity),而急迫性是在国际人道法允许的范围之内,结合具体的作战背景、条件进行的复杂判断。自主武器系统对是否符合比例原则的判断,比区分原则更为艰难。比例原则要求针对攻击方预期获得的军事利益与可能产生的附带损害之间取得平衡,但在敌对行动中,军事利益与附带损害并不单纯地表现为数量比值关系。采用在自主武器系统中设定参数、建模、概率测算方法,将问题转化为数学运算,对比例原则的适用是有疑问的。军事操作层面可以采用“附带损害评估方法”(Collateral Damage Estimate Methodology)规范自主武器使用,将自主武器此方面能力不足转为人的决策,^⑩但这一操作机制本质上属于对人的评估和授权问题,运用这一机制意味着战场指挥官辅助自主武器在回路中决策,实际上反映了自主武器智能化程度不足以独立判断攻击是否符合比例原则要求。

预防措施原则在适用中也面临障碍。传统上,预先警告由发起攻击方的战斗员完成,该原则适用于自主武器场景的障碍在于武器的开发思路和实际操作,即自主武器中是否加入预先警告功能受到武器使用场景、性能指标、操作特性等一系列因素的制约,并非受法定情况下(如平民与战斗员混杂)发起攻击方必须履行警告义务的制约,因此自主武器系统并非必须带有这一功能。

此外,国际人道法规则经常面临叠加不同法律规则并经综合考虑后,才能决定适用与否的判断。例如《第一附加议定书》第56条,默认攻击含有危险力量物体是违法行为。例外情况是,若该物体被用于其通常作用以外的目的,和用以使军事行动得到经常、重要和直接支持的目的,且如果这种攻击是终止这种支持的唯一可能方法,则允许进行攻击。除此之外,在该议定书第36条第3款下,对环境造成广

泛、长期、严重影响作战方式方法也应被禁止,因此,对发电厂、水坝等包含危险力量物体的攻击,本身包含复杂的规则叠加情形。在人工智能领域,算法尚未能帮助自主武器系统具备驾驭上述规则必需的认知能力和判断能力。

综上所述,操作员行为与自主武器系统攻击的不一致性,使得武器的攻击动作并不能在所有情况下都如实反映操作员及指挥官的真实意图,也不能保障自主武器的每一次攻击都是受控的。有意义的人类控制原则正是在这一背景下被提出,既然自主化乃至智能化意味着不确定性,特别是战争权力委托造成武器在人工智能深度运用下存在不可控风险,因此各国发展、使用致命性自主武器应尽最大可能,使武器系统保持在国家及其武装部队的控制之下。

二、有意义的人类控制原则:共识及理解思路

CCW框架下,有意义的人类控制原则被引入2017年GGE会议讨论。2018年,GGE会议报告将有意义的人类控制列为未来谈判方向和10项应遵循基本原则之一。^⑪对有意义的人类控制原则一开始并没有十分详细的诠释,但主要军事大国基本对此存在一定程度共识:例如美国国防部在2012年颁布3000.09号指令,用以专门规制自主武器系统的开发、测试、部署和使用,该指令将自主武器系统置于人类控制之下,并使自主武器的使用能充分反映战地指挥官和操作员的意图,要求自主武器系统的设计应该允许指挥官或操作员以适当的方式对使用武力作出人类的判断。^⑫俄罗斯在2017年的GGE会议中表示,尽管有意义的人类控制原则尚未被完全挖掘含义,但这一原则应当继续和发展。^⑬2018年GGE会议中国提交立场文件,对致命性自主武器系统的特征描述包括具有自学习能力、一经激活不能关闭等、具有完全自主性的武器系统。^⑭中国建议,国际社会应该就禁止开发和使用致命性自主武器达成简短的国际文书。这意味着,中国认为应对致命性自主武器设定较高的门槛条件,并禁止研发和使用越过这一门槛条件的武器系统。可见,中国也认为武器的自主性应受到监管。日本在2019年GGE会议提交的工作文件确认,该原则是在武器造成违法攻击的情况下追究责任的前提,因

此有必要采取措施防止武器系统脱离人工控制;此外,有意义的人工控制的各种表现形式,有待于各利益相关方进一步讨论。²¹

通过对各国军控法规、政策酝酿及GGE会议文件汇总,有意义的人类控制原则应表达为在自主武器的政策制定、开发、部署、测试、使用、使用后评估中,尽最大努力贯彻人对武器系统的控制。2018年GGE会议报告,将上述过程用数字“0”至“5”予以标注并绘制成图表(日出图),方便各国理解并调整政策。²²此后各国发布政策,多在日出图提供的框架之内。

日出图提供了有意义人工控制原则的一种理解方式,落实该原则的着眼点是通过国家自律的行为来增强对自主武器控制的确定性,对武器开发至使用的全流程控制。这种视角下,对军控义务的探索属于初级规则层面的讨论,未来的表现方式将是围绕武器开发至使用的全流程,对各国行为提出要求。

理解有意义的人类控制原则的第二种思路,是从法律责任角度切入。2018年GGE最后报告指出,不应把违法攻击导致的责任转移给机器²³,意即武器违法攻击的责任,始终应由个人或国家承担,其对应的是战争罪引发的个人刑事责任和国家责任。事实上,第二种思路早在2017年的GGE会议报告中就形成了结论。²⁴上述两种理解思路相辅相成,完成了法律义务到法律责任的闭合。截至2019年,GGE对有意义的人类控制原则的讨论多集中在日出图思路,其原因是CCW提供的法律框架,军控的落脚点

始终是武器的禁止、限制使用及周边问题如透明度、防扩散等,而使用后的责任问题属于国家责任法和国际刑法(追究个人刑事责任)的范畴,次级规则自成一体,有自身的运行逻辑。

三、有意义人类控制原则和对国家行为的要求

(一)全流程监管及法律问题

2018年GGE报告把全流程监管划分为由数字“0”至“5”标定的6个阶段,分别对应:(0)研发前政策指导;(1)研发;(2)测试、评估和认证;(3)部署、训练、指挥和控制;(4)使用和放弃;(5)使用后评估。具体表示如下:

对于国家而言,全流程监管意味着所有阶段均由国家的法律法规控制,技术化规范对应研发阶段及测试、评估、认证阶段涉及产研标准问题;既有国际法规则体现在军控规则和国际人道法方面,因此第三、四阶段受到国际法和国内法的双重制约。按照日出图的逻辑,第三、四阶段的违法性,国际法上和国内法上的义务和责任互不影响。还需明确,1977年《第一附加议定书》第36条规定,在研究、发展、装备及使用新武器和新作战方式、方法前,缔约国有义务审查其是否被议定书和其他有效的国际条约禁止(本文以下简称为“第36条审查”)。从日出图看,第36条审查可能存在于两个阶段,第一种可能性是第二阶段中对武器的测试、评估和认证中,包含了法律意义上的审查,缔约国履行第36条审查义务纳入第二阶段中;第二种可能性是在第三阶段,把第36条审查作为独立程序。不论哪种可能性,第36条审查都是缔约国应尽的义务。

已经表明立场或发表意见的国家已就全流程贯彻有意义的人类控制取得了基本共识,但对细节问题的理解存在一定程度的差异,这些差异可能成为未来谈判的走向因素。

第一,如何理解第36条审查的问题。代表性国家都强调第36条审查的重要性及履行审查义务意愿,但都不明确支持建立具有强制约束力的多边机制,甚至反对建立多边机制。²⁵在已表达立场的国家中,日本的观点比较特殊。日本认为应在

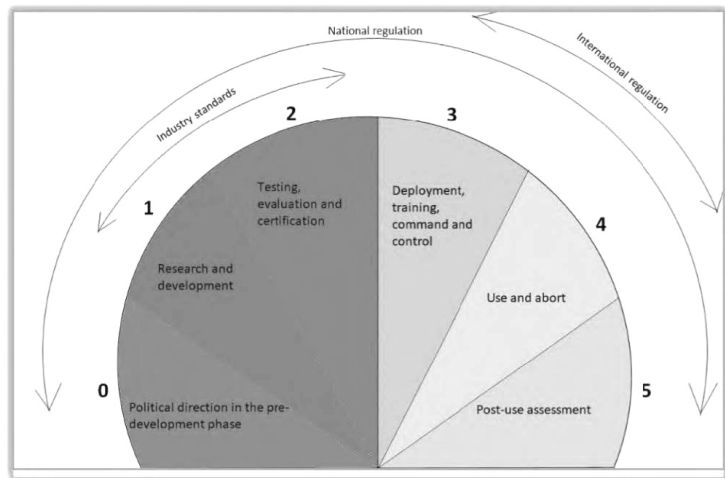


图1 日出图(摘录自2018年GGE会议报告)

CCW框架下加强信息共享、透明度建设,探索可被一致通过的机制,根据《第一附加议定书》第36条的规定,引入缔约国如何遵守该条要求的机制,可以构成一个选项。^③此方式实际上意为在多边军控机制(条约)下,将自主武器系统的法律审查作为统一标准义务。按照这种理解,第36条审查已经不仅仅是《第一附加议定书》下的义务,而应该被未来拟议中规制自主武器的专门军控条约涵盖,此种观点过于超前。根据约文通常含义,第36条审查是缔约国自行完成的义务,新武器和与之对应的新作战方式方法是否被禁止使用,依赖于缔约国自行判断。根据红十字国际委员会解释,约文原意是,第36条仅要求缔约国对新武器通常使用范围是否符合国际人道法要求进行审查,^④第36条并没有要求在审查国与其他缔约国之间建立一种强制约束力机制,也不要求审查国向任何第三国公开审查结果。^⑤究其原因,第36条仅规制缔约国对新武器的使用,而非回答拥有新武器是否合法的问题,后者属于裁军法律规则的调整范围。^⑥由此可见,《第一附加议定书》下平衡各方利益的最佳方式是,缔约国在第36条下有义务对新出现的自主武器进行审查,但无义务对任何第三国公开。在军控法规体系下,未来规则强制性成分的高低,直接决定了军事大国受到未来自主武器军控规则制约的意愿。强制性过高导致军事大国不愿意受到约束(如俄罗斯明确反对将第36条审查义务作为强制性义务引入CCW的多边框架^⑦),将会产生类似核大国不愿签署和批准《禁止使用核武器条约》的困境。因此在CCW框架下将第36条义务多边化、强制化,对有能力开发和部署、使用自主武器的军事大国显然不利。

第二,人机交互界面(human-machine interface)设计的问题。与传统武器不同,自主武器实现自身功能依赖武器系统从人机交互界面获得正确的输入信息,若交互界面作为输入端不能给自主武器提供正确的信息,武器系统工作反映操作员输入的攻击意图就得不到保证。因此,人机交互界面是同时牵涉武器设计、生产以及使用者各层面问题,也是法律问题。例如,美国防部3000.09指令要求,设计自主武器系统必须考虑未经计划的交战或者丧失武器系

统控制的风险,软硬件应包含安全机制、抗干扰机制、信息保证机制等;人机交互界面应达到严格的要求;能够为接受过相应训练的操作员了解、提供对系统状态的可追溯的信息反馈机制、为操作员激活或关闭武器系统功能提供清晰的程序。^⑧综上所述,人机交互界面问题在理论上简明易懂,但实现需要由武器设计、测试和作战人员训练等多方面保证才能达到结果:一是武器设计的可靠性论证;二是人机交互界面及背后算法可靠性测试;三是交互界面及操作流程能否为指挥官、操作员理解并排除误操作可能性;四是武器列装后,指挥官、操作员根据自主武器系统的应用场景、应用规范持续训练。

第三,自主武器测试、评估和认证问题。各国武器开发流程不同,但武器性能及可靠性是否达标是评价其能否得到装备的必要条件。测试、评估及认证,是正式装备前的必要流程,自主武器对可靠性测试提出了更高的要求。美国立法规定,不仅要对武器系统的整体进行测试,还要对各子系统的可靠性进行测试;对武器性能做出任何改变,必须对整个系统进行可靠性测试。^⑨

第四,使用自主武器的范围限制和交战规则的问题,即对自主武器的使用范围以及交战规则作出明确规定,以此限制违法使用的可能性。例如澳大利亚强调,自主武器应当确定其使用边界,即确定在何时、何地以及用何种方式使用的限制条件,^⑩以及在军事行动中,指挥官、操作员使用自主武器的,应该遵守相应交战规则的限制。^⑪又如在美国的法规中,使用自主武器分为法律授权范围内和授权范围外使用两种情况,授权范围由3000.09号指令明确规定。超出授权范围的使用以审批为前提条件,须经主管政策的国防部副部长、主管装备、技术和后勤的国防部副部长,参联会主席批准。^⑫

(二)防扩散机制

除了全流程监管,自主武器还存在防扩散的问题。这一问题的产生,来自人工智能技术易于从武器系统中剥离的特点,即承载了自主武器技术的算法、软件等,可以独立于硬件实体成为流通标的。2012年《武器贸易条约》对禁止、限制武器贸易设定了一系列条件,但列入条件的规制对象仅仅包括常

规武器及其弹药和零部件(第2、3、4条),而驱动武器并实现武器自主化的软件、算法并不在此范围之内,但自主武器技术的扩散和滥用同样可能造成严重的人道主义危机。从应然法的角度出发,对实现武器自主化技术进行管制并建立适当的终端控制机制,应当被纳入考虑范围。

四、有意义的人类控制原则和责任规则

以全流程监管思维来看,从国家责任和个人刑事责任层面对违法使用武器的追责依据,属于初级规则的自然延伸。在操作员行为与自主武器保持一致性的情况下,自主武器攻击若未违反国际人道法,追究国家责任和战争罪(满足主客观要件)并不存在问题,理论上的难点主要在操作员行为与自主武器攻击不一致的情况。在使用流程中,操作员通过人机交互界面完成武器攻击设定,但正如前文所言,法律的模糊性、战场物理条件限制及算法本身的问题,不能排除武器得到交互界面设定后能够反映操作员的意图。因此,超出人机交互界面设定、由武器系统“基于自主性发动的攻击”(以下简称“超设定攻击”),如何确定操作员的个人刑事责任和国家责任,是探讨的难点。

(一)个人刑事责任追究思路及其障碍

法国政府及某些学者的观点支持运用指挥官责任解决个人刑事责任问题。^⑧这一观点实际上是将自主武器系统拟制为法律人格,为了追究指挥官个人刑事责任而套用指挥官责任,是虚拟人格思路的反映。但严格考察指挥官责任的构成要件,会发现将其用于解释超设定攻击引发的战争罪问题存在障碍:(1)该设定面临上下级关系悖论。习惯法和国际刑事法院的法理均确认判断上下级关系的法律标准是指挥官本人能否对下级进行有效控制,因为有效控制是指挥官预防或惩治下级犯罪现实能力的前提。^⑨这一前提,自协约国为缔结《凡尔赛条约》成立的战争发动者及执行惩罚委员会提出追究指挥官责任不作为责任的观点起,就得以强调。^⑩在超设定攻击条件下,指挥官若能有效控制自主武器这个拟制的“下级”,就意味着不存在自主武器突破操作员设定实施违法攻击的可能性。(2)偶发性的超设定攻击来源于技术障碍,即便指挥官尽职尽责履行注意义务也

可能发现不了,因此不满足指挥官责任构成要件中知道或应当知道下级实施犯罪的要求。(3)《罗马规约》第28条规定,指挥官的不作为行为与下级实施犯罪之间的因果关系是指指挥官责任的构成要件,国际刑事法院将其解释为指挥官不作为增加了下级实施犯罪的风险。^⑪在自主武器超设定攻击的场景特别是偶发的超设定攻击场景下,指挥官的不作为行为与攻击之间只有相关性,而非因果性。

与拟制主体思路相对应的是超设定攻击工具论,即犯罪工具的动作在多大程度上超出了犯罪人的预先判断。哥特维纳等人案(Prosecutor v. Gotovina et al.)中,法庭暗示,误差可能不在被告人积极追求范围之内,但可以在犯罪人预料的范围之内。在国际刑法习惯法规则中,战争罪的构成要件包含间接故意,超设定攻击在操作员预见范围内的,可认定为间接故意的主观状态,由此可在一定程度上化解自主性带来的个人刑事责任难题。值得注意的是,战争罪的主观要件是否包含间接故意这一问题,习惯法和条约规则存在不同认识。虽然根据前南斯拉夫国际刑事法庭(以下简称“ICTY”)“对严重违反日内瓦公约行为”(ICTY 规约第2条)规定,被告人主观心态包含轻率(recklessness)状态,^⑫但由于法庭不追究过失责任,因此在ICTY的法理中,习惯法上的“轻率”等同于大陆法系的“间接故意”。然而,在国际刑事法院的体系下,《罗马规约》第30条所指的“故意”仅指直接故意;法院不支持规约规定战争罪主观要件包含间接故意的观点(在与使用武器且与“攻击”有关的罪名中,国际刑事法院针对三类行为发表了不包含间接故意的观点,分别为故意杀人^⑬、摧毁或毁损宗教场所^⑭、军事必要非法且肆意摧毁财产^⑮)。上述情况导致的技术性问题是,采用不同的国际刑法规范体系追究超设定攻击引起的个人刑事责任,将存在不同的逻辑结果。当然,习惯法和《罗马规约》并不禁止各国国内法以构成间接故意为主观构成要件,起诉、判决战争罪行为。

(二)国家责任:第8条路径及理论障碍

按照国际法委员会二读的《关于国家对国际不法行为的责任条款草案》(以下简称《国家责任条款草案》)的观点,主观状态不构成国家责任的构成

要件。因此,可归因性对超设定攻击不法行为的包容性问题,理论上存在几种主张:一是以第8条(指挥控制非国家实体)为依据实现归因;二是以《国家责任条款草案》第4条(国家及其机构自身行为)为依据实现归因;三是以国家承担预防(监管)义务为前提,以超设定攻击情况下国家的不作为违反预防义务作为追责基础。第三种路径对应的是《国家责任条款草案》(一读)的第23条,然而不论考察何种依据,均存在理论障碍,从而影响有意义的人类控制原则的贯彻。

《国家责任条款草案》(二读)第8条描述的是国家指令(instruction)、指挥(direction)或控制(control)非国家实体(个人或团体,如反政府武装),从而为非国家实体违法行为承担责任的情形。詹姆斯·克劳福德(James Crawford)起草《国家责任条款草案》之时,ICTY的判例尚在发展中^⑧,因此文本中“指令、指挥或控制”的法律标准根据第8条的含义无从确定。司法实践中分为两个流派,一是严格遵循国际法院在尼加拉瓜案(Nicaragua v. United States of America)中建立的有效控制标准(effective control test),即只有明确找到一国政府向非国家实体发出特殊指令(specific instruction)的直接或间接证据时,才能判定非国家实体成为国家的代理人或工具,从而将非国家实体行为归因于国家成为可能^⑨;二是ICTY在塔迪奇案(Prosecutor v. Duško Tadić)上诉判决中确立的全面控制标准(overall control test)。全面控制标准讨论的是南联盟共和国政府是否控制了波黑共和国境内的塞族反政府武装,从而使得波黑境内非国际性武装冲突演变为国际性武装冲突。虽然这是一个武装冲突性质判断问题,但ICTY上诉庭在判决中阐明,判断的逻辑起点是国家责任规则中对非国家实体的归因问题。^⑩ICTY认为,应当区分单个个人和有组织团体及个人同化并在国家机构的体系下行事三种情形的讨论:^⑪对于前者与国家的控制关系,应当继续适用国际法院的有效控制标准;对于有组织性的团体如反政府武装部队,应当适用全面控制标准,全面控制标准在证据法层面的表现就是,法庭并不要求一定要找到特殊指令的证据,而只需证明国家对非国家实体的控制达到相当高的程度即可^⑫;国际法院在

《防止和惩治种族灭绝罪公约》适用问题案(Case Concerning the Application of Genocide Convention)中指出,判断国家责任的归因性问题,应采取有效控制标准,判断武装冲突性质采用全面控制标准。^⑬但国际法院对于为何作出上述区分,以及为何要避免对ICTY关于考虑判断标准的逻辑起点是归因性论述作出直接评判,都没有给出解释。国际法院判决的惯性导致国际刑事法院采取了上述妥协的立场。^⑭

ICTY上诉庭没有阐述全面控制标准的前提是应用于人的组织还是应用于拟制为武装冲突参与方的武器系统,但波黑战争的背景不涉及人工智能技术的问题,因此全面控制标准适用于国家与非国家实体的关系是显而易见的。从上诉庭所评估的事实看,自主武器与一个具有高度组织性和纪律性的非国家实体性质不同。在实务操作中,采用有效控制标准必须证明国家对作为非国家实体的组织发布了特殊指令,但是笔者认为,就规则适用的前提条件来说,全面控制与有效控制标准并没有不同。在国际法院确立依据的两个案件中,一是讨论尼加拉瓜反政府武装力量(contra force)与美国政府之间的关联问题,二是在《种族灭绝罪公约》适用问题案中讨论南联盟共和国(塞尔维亚)是否应对波黑塞族的“塞普斯加共和国”军队实行种族灭绝罪负责的问题。就适用前提而言,国际法院与ICTY讨论的情况完全相同。近年来,司法或仲裁案件中适用《条款草案》第8条的主要案件情况如下页表1所示。

由表1统计可见:(1)第8条的适用存在两种主要情形:一是反政府武装组织被一国政府指挥或控制的情况,本条用于解释控制国政府是否对违反国际人道法的行为负责;二是由国家控股或作为股东的公司在何种情况下被视为国家代理人,从而将公司违反投资法(同上是双边投资条约)义务的行为归因于国家。(2)两种非国家实体自身由人组成,决策、行动体现人的意志自由和行动自由。(3)国家致力于建立非国家实体的事实,不能作为可归因性存在的决定性证据,接受程度最高的“有效控制”标准下,国家向非国家实体发出特定指令才是建立可归因性的关键证据。同理,仅国家部署自主武器系统的事实不能单独作为可归因性的证据。(4)自1986年尼加拉

表 1

	讨论关系	审理机构	支持标准或观点	属性
Kenneth P.Yeager	伊朗政府与革命卫队关系 [®]	美伊求偿法庭	全面控制标准	被 ICTY 引用法律依据
Loizidou v. Turkey	“北塞浦路斯共和国”与土耳其政府关系 [®]	欧洲人权法院	全面控制标准	被 ICTY 引用法律依据
Jorgić case	波黑战争期间围攻萨拉热窝的波黑塞族部队与南联盟共和国的关系 [®]	德国杜塞尔多夫高等法院	全面控制标准	被 ICTY 引用法律依据
Prosecutor v. Duško Tadić	波黑战争期间波黑塞族部队与南联盟共和国的关系	ICTY	全面控制标准	ICTY 由此案确立全面控制标准
Nicaragua v.US(尼加拉瓜案)	尼加拉瓜境内不明拉丁美洲国家人员组成的武装力量(contra force)与美国政府关系	国际法院	有效控制标准	国际法院未说明“有效控制”标准的法律依据
Case Concerning the Application of Genocide Convention(《防止和惩治种族灭绝罪公约》适用问题案)	波黑战争期间波黑塞族部队与南联盟共和国的关系	国际法院	有效控制标准	引用尼加拉瓜案为依据
Prosecutor v.Lubanga	刚果民主共和国境内反政府组织(刚果爱国者联盟,UPC)是否受邻国乌干达控制 [®]	国际刑事法院	有效控制标准(承认全面控制标准适用于判断武装冲突性质)	引用 Tadić 案
Prosecutor v.Katanga	刚果民主共和国境内反政府组织(刚果爱国者联盟,UPC)是否受邻国卢旺达控制 [®]	国际刑事法院	有效控制标准(承认全面控制标准适用于判断武装冲突性质)	引用 Tadić 案
Encana Corporation v. Republic of Ecuador	厄瓜多尔石油公司作为一个国有企业是否构成由国家控制的私人实体 [®]	伦敦国际仲裁院	未展开讨论	引用《条款草案》第 8 条
United States—Countervailing Duty Investigation on Dynamic Random Access Memory Semiconductors(DRAMS)from Korea	私人公司行为在何种情况下被国家控制 [®]	WTO 上诉机构	未展开讨论	引用《条款草案》第 8 条
Ceskoslovenska Obchodni Banka A.S.v.The Slovak Republic	半国有半私营公司在何种情况下对其他公司的投资建议被视为政府行为,从而构成对双边投资条约的违反	ICSID	将半官方半私有公司 SODIGA 的行为分为行政属性和商业属性,商业属性的行为不得归因于国家,而判断履行政府职能的行为的属性,其决定因素是行为的性质而不是行为的目的 [®]	同时引用《条款草案》第 5、8 条
Emilio Agustín Maffezini v. The Kingdom of Spain	同上	ICSID	标准同上 [®]	同上
Beijing Urban Construction Group Co. Ltd v. Republic of Yemen	同上	ICSID	标准同上 [®]	同上
Saint-Gobain Performance Plastics Europe v. Bolivarian Republic of Venezuela	半国有半私营公司在何种情况下被视为委内瑞拉总统指挥或控制	ICSID	仅在事实上听从查韦斯总统指挥、指令或被控制时才满足第 8 条要求 [®]	引用《条款草案》第 8 条
Ampal- American Israel Corporation and others v. Arab Republic of Egypt	埃及天然气控股公司终止《天然气销售协议》的行为在何种程度上归因于埃及政府	ICSID	标准同上 [®]	同上
Hulley Enterprises Limited(Cyprus) v. The Russian Federation	国家建立公司行为本身是否说明公司的行为可归因于国家	常设国际仲裁院	国家建立公司行为不构成将公司行为归因于国家的基础,归因性来自国家对公司的控制、指挥 [®]	同上
Marfin Investment Group Holding S.A., – Alexandros Bakatselos and Others v. Republic of Cyprus	国家占有公司一定股份并任命执行董事的事实,是否表明公司行为归因于国家	ICSID	塞浦路斯政府在公司中持有股份的事实,不足以将公司的行为归因于塞浦路斯政府 [®]	引用第 8 条及国际法院观点

瓜案以来,第8条从未用于直接解释机器与国家之间的关系,即第8条的适用场景中从未出现非国家实体为拟制法律人格者的情况。综上,将第8条用于解释超设定攻击违反国际人道法的行为,实际上是法律的类比适用。笔者认为,能否类比适用取决于自主武器的自主性与非国家实体的意志、行动自由是否存在同质性或者相似性。以塔迪奇案中卡拉季奇领导的波黑反政府政权“塞尔普斯加共和国”(Republika Srpska)及其军事力量(VRS)与南联盟共和国之间的关系进行详细比较,如表2所示,事实因素表明,“塞尔普斯加共和国”作为由人组成、受控制或指挥的组织,具有相当的自我组织能力、决策能力。

笔者认为,结合表1、表2,适用于表征国家与组织之间的因素,不能移植到“国家—机器”关系的分析中。国际法院认为“有效控制”是在《条款草案》第8条下判断国家责任的唯一标准,采用此标准将带来控制有效性的悖论:只有国家发出特定指令,第8条的适用条件才得到满足,但超设定攻击的出现,恰恰是自主性导致武器的攻击动作违背人机交互界面中操作员输入指令的结果。

(三)国家责任:第4条路径及理论障碍

适用《国家责任条款草案》第4条解释自主武器超设定攻击的归因性,分析的框架是“国家—武装部队(操作员)—机器”三者之间的关系,回答问题的本质是超设定程度在何种程度、范围内,被视为国家机构行为的延伸。自主武器系统不被赋予拟制的法律人格,仅作为犯罪工具来处置,其归因逻辑是间接式的,可分解为两个步骤:第一步,超设定攻击违反国

际人道法,是否构成操作员行为的延伸;第二步,操作员作为武装部队的成员,行为是否当然归因于国家。理论上的难点不在第二步而在第一步是否成立,这是国际刑法中判断个人刑事责任的问题,须将国际刑法中正犯主观要件的认定嵌套到归因性问题中。按照这一逻辑,国际刑法应对超设定攻击的不足,也构成依照第4条追究国家责任的不足,对于操作员根本无法预见的超设定攻击,国家不对此负责。

采用第4条作为解释国家责任的基础,意味着国家责任的成立取决于武装部队成员的刑事责任是否成立。是否存在一种路径,将超设定攻击导致的个人刑事责任和国家责任分开处置、互不干扰呢?历史上仅有一起先例,即国际法院讨论波黑塞族部队中众多被告人成立种族灭绝罪,同时南联盟共和国(塞尔维亚)因为控制或指挥塞族部队同时成立国家责任的可能性的例子。国际法院在《种族灭绝罪公约》适用问题案(Case Concerning the Application of Genocide Convention)中认为,个人刑事责任和国家责任是并行不悖的双重责任体系。^⑥国际法院并没有分析二者之间的关联性,二者在法律适用上也甚少交叉。在国际法院看来,二者的关联性可能来自证明标准的趋同:追究可归因于国家的种族灭绝行为,属于有极端危害性(exceptional gravity)的国家不法行为的问题,优势证据标准在此不适用;国际法院对用来证明违法行为的证据具有充分的决定性,以使证据有更高程度的确定性,该证据的确定性程度与争议行为的严重性相称。^⑦国际法院的方法是将个人刑事责任与国家责任的关系演化为证据法问题,使国

表2

比较项目	Republika Srpska	自主武器系统
动机	割据波黑部分领土并独立为塞尔维亚族国家,或与南联盟共和国合并为更大的塞尔维亚族国家。	无动机
目标	将塞族区域境内的穆斯林族永久驱逐或杀害。	无目标
组织结构	完整的武装部队组织体系。	自主武器的控制结构
指挥体系	来自南联盟共和国军队建制部队的军官被移植,指挥体系完整。	指挥官—操作员—自主武器
财政能力	南联盟共和国支付军官收入,并建立若干人员服务中心,为在波黑塞族军队工作的军官提供服务。	无财政能力
外部支持	南联盟共和国提供各种物资、武器、弹药支持。	使用国的C4ISR(指挥、控制、通信、计算机、情报、监视、侦查)体系为获取目标数据提供支持
自由意志及意思表示能力	作为参与方参加保证波黑战争各方遵守最低限度国际人道法要求的《代顿协议》的谈判案。	无意思表示

家责任的证明标准与刑事责任证明标准趋同,以便解决二者潜在的冲突。这一做法不无道理,但仅能运用于可预见的超设定攻击带来的责任问题,对于操作员不可预见超设定攻击,问题的本质是实体法并没有提供合适的路径,并不是调整证明标准就能解决的。

(四)国家责任:以不作为违反预防(监管)义务追责及理论障碍

GGE会议对有意义的人类控制原则的讨论表明,为国家设定监管自主武器并预防自主武器不受控的违法攻击,有可能成为未来法律规则的内容。不同国家的法规和政策均表达,自主武器使用中操作员、指挥官应持审慎态度的立场,这也是从逻辑上落实有意义的人类原则的客观要求。因此,在未来存在以国家负监管自主武器、防止其不受控的义务为前提,以国家的武装部队的不作为为依据追究责任的路径。科孚海峡案(The Corfu Channel Case)中,国际法院面对阿尔巴尼亚是否对英国军舰通过科孚海峡触发水雷负责的问题中,国际法院肯定了谨慎注意义务:判决认为,国家不得在明知的情况下,允许自己的领土被利用去实施有悖于其他国家权利的行为。^⑥刚果领土军事行动案中,乌干达军队占领刚果(民)伊图利地区(Ituri District)期间是否应当对被占领土的资源被不属于乌干达部队的私人组织劫掠的行为负责的问题,国际法院从谨慎注意的角度出发,认为乌干达作为占领方,其不作为行为违反《海牙陆战章程》第43条的规定。^⑦可见,由不作为违反谨慎注意义务而导致国家责任,规避了第4条、第8条适用中的障碍,将核心问题转换为国家的不作为对谨慎注意义务的违反。

尽管如此,以不作为作为追责依据仍然存在一系列难题:

第一,先例适用谨慎注意义务都存在前提,即相关领土处于国家的控制或长期关注之下。特雷尔冶炼厂案(Trail Smelter Case)中,污染源来自加拿大领土内;科孚海峡案(The Corfu Channel Case)中,英国军舰触雷地点长期处于阿尔巴尼亚的观察和监控之下^⑧。而在刚果领土军事行动案(Case Concerning Armed Activities on the Territory of the Congo)中,虽然

事发地点在刚果境内,但该地处于被乌干达政府占领状态。^⑨在国家与自主武器系统的关系中,自主武器的超设定攻击可能是偶发,部署自主武器的地点也不限于一国的领土或其有效控制的领土,因此不能假设国家的武装部队对自主武器系统在偶发的超设定攻击之前就必须保持极高程度的关注。与讨论情形最接近的是2007年国际法院在《种族灭绝罪公约》适用问题案(Case Concerning the Application of Genocide Convention)中面临的状况:针对波黑斯雷布列尼察的种族灭绝行为,国际法院肯定塞尔维亚在《防止和惩治种族灭绝罪公约》下有义务预防种族灭绝罪,但塞尔维亚仅在种族灭绝罪被开始实行的情况下才承担此项义务。^⑩该结论的依据并非来自《种族灭绝罪公约》下的特定规则,而是基于《国家责任条款草案》第14条第3款的规定,^⑪因此具有习惯法上的普遍意义。国际法院由此认为,预防义务并不是种族灭绝罪开始实行,或者被试图开始实行的时候就产生的,如此推论是荒谬的。事实上,国家的预防义务以及采取相应行动的义务,应在国家知道或正常情况下应当知道种族灭绝罪的实行带来“严重风险”的情况下才产生。^⑫斯雷布列尼察种族灭绝的情况与国家监管自主武器的情形完全不同。前者背景下,种族灭绝行为是在特定时间段内有南联盟共和国支持的波黑塞族武装力量反复大规模实行的,并且持续了相当时间;但在后者背景下,超设定攻击违法是偶发的,在操作员保留对回路控制权的情况下,可关闭武器系统防止超设定攻击的反复出现。

第二,预防义务的性质将影响责任承担。预防义务的条款最初出现在国际法委员会1996年《国家责任条款草案》(一读)之中。对于国家的不作为,该一读草案还有不作为违反的是行为义务(obligation of conduct)和结果义务(obligation of result)。即行为义务要求国家做到特定行为,而结果义务的遵守是国家保证违反义务的后果不能出现。按照克劳福德的解释,二者区分来自大陆法系的传统,行为义务也被称为“最大努力义务”(best efforts obligation),只要国家做到了初级规则要求的行为,就不算作国际不法行为。^⑬同时,国际法上的义务被定为行为义务还是结果义务,由初级规则而非规制国家责任的次级规

则决定。^⑤行为义务与结果义务的分类,并没有体现在《国家责任条款草案》(二读)之中,但国际法院在《种族灭绝罪公约》适用问题案(Case Concerning the Application of Genocide Convention)中认可了分类,并认为公约下的预防义务具有行为义务的属性。^⑥行为义务与结果义务的差异及国家责任的类型提醒我们,未来若针对自主武器的监管形成具有强制约束力的国际法规则,其被界定为行为义务还是结果义务,对军事技术大国来说至关重要。军事力量的强弱可能再次预设了力量,对军事强国而言,监管义务被界定为行为义务更有利。

第三,按照既有国家责任规则,国家就其不作为违反预防义务承担责任的情况,应当区别于转承责任或间接责任。转承责任的理论最早来源于奥本海的倡导,^⑦但该理论没有得到广泛支持。近年来网络攻击法律问题的兴起,导致转承责任理论受到重视,例如美国外交和领事人员案(Case concerning US Diplomatic and Consular Staff, 又称 US v. Iran)中,被解读为伊朗在明知情况下为替代伊朗民众承担责任的依据,以此映射网络攻击情况下,国家应对其境内的个人、组织攻击另一国的行为承担责任。^⑧实际情况并非如此。国际法院将伊朗政府对德黑兰民众的行为划分为两个阶段:第一阶段的民众自发行为本身不能归因于伊朗,但第二阶段在霍梅尼宣布占领美国使领馆并将外交、领事人员劫持人质的政策以后,“这项政策的结果从根本上改变了占领大使馆和将其外交和领事人员作为人质拘留造成的局势的法律性质。阿亚图拉·霍梅尼和伊朗国家其他机构对这些事实的批准,以及使这些事实永久化的决定,导致了对大使馆的持续占领和劫持人质转变为国家的行为”^⑨。国际法院的认识中,伊朗政府对不法行为的“批准”决定,导致区分责任是从第一阶段占领美国使领馆开始起算还是从伊朗“批准”之时起算的问题没有意义,国际法院认为,伊朗对此第一阶段的不法行为负责的法律基础是国家自身的接受,而非国家怠于采取足够的措施从而防止占领使领馆和劫持人质事件的发生^⑩。依照上述路径,转承责任并没有被国际法院接受。国际法委员会是在《国家责任条款草案》第11条(国家对不法行为的事后接受、确认)的

基础上理解德黑兰人质事件中伊朗的责任,而非转承责任。^⑪回到国家不作为违反预防自主武器超设定攻击义务的命题之中,并不存在武器系统“自主”决定的超设定攻击构成违反国际人道法,国家需要“替代”武器系统这一虚拟的个人主体承担责任的情况。

第四,国家不作为与自主武器超设定攻击之间的因果关系不能建立,在超设定攻击违反国际人道法属偶发状况的情况下尤其如此。对于不作为违反预防义务引发的国家责任,因果关系评价问题可追溯到1930年海牙国际法编纂会议筹备委员会向各国征求意见的第8项,即个人行为致使外国人受到侵害的情况下国家未能履行预防义务导致国家责任的内容。奥地利对此回复认为,仅仅国家的不作为而未能履行预防义务本身不能构成国家责任,损害必须考虑因果关系。起草委员会吸收该意见形成了起草基础事项第10项,对此予以确认。^⑫此后,在罗伯特·阿戈担任国际法委员会国家责任议题特别报告员期间,因果关系被再次强调。^⑬《国家责任条款草案》(一读)第23条下,确认了上述框架。国家的不作为、应当被预防外部事件(external event)以及二者之间的因果关系缺一不可。^⑭此后,《国家责任条款草案》(二读)把因果关系排除在国家责任的构成要件之外。根据克劳福德解释,国际法委员会认为因果关系应当由初级规则决定,而非次级规则决定。^⑮因此,对国际法委员会立场更确切的表述是:若初级规则包含因果关系,那么国家责任的构成要件也应包含因果关系;若初级规则不包含因果关系,那么因果关系不影响国家责任成立,因果关系是构成损害赔偿的前提。CCW框架下拟定军控规则,是否包含违法的因果关系是不可预料之数,目前尚未有国家在GGE会议发表关于这一问题的观点和立场。若按照《国家责任条款草案》的默认规则理解,国家的不作为与超设定攻击造成的损害结果之间很难建立因果关系,其原因来自两方面:一方面是技术因素,超设定攻击产生源于武器自主性带来的人的行为和武器攻击之间不一致性,复杂作战背景放大了这一不一致性,换言之,不论操作员、指挥官尽多少谨慎的注意,超设定攻击总有可能发生;另一方面,国际司法、

仲裁对因果关系的认定总是持保守态度,按照郑斌教授对早期司法和仲裁实践的总结,损害赔偿中因果关系属于一般法律原则的范畴,近因原则占据主导地位。^⑥在国家的不作为违反行为义务的情况下,如何界定不作为与损害结果之间的因果关系更是难题。以波黑境内的斯雷布列尼察种族屠杀为例,国际法院确认南联盟共和国违反了预防种族灭绝的义务,但以波黑没有证明南联盟共和国的不作为与斯雷布列尼察发生的种族灭绝之间存在足够的联系为由,拒绝满足波黑赔偿的请求。^⑦国际法院对因果关系的要求是,“如果被告根据其法律义务行使(尽到了预防义务),那么斯雷布列尼察的种族灭绝在事实上可以避免”。^⑧由此可见,国际法院采用的是“如果没有……就不会……”的公式(but for test)或必要条件说(conditio sine qua non)。采用以上方法判断超设定攻击违反国际人道法的情形,实际上根本不能解决赔偿的问题。另一个值得注意的动向是,在国际人道法领域,国际司法实践引入了可预见性标准,用以确定损害赔偿及其范围,即赔偿与否及赔偿多少以不法行为一方事前可合理预见范围为限度。^⑨然而,超设定攻击超出操作员、指挥官的预料是当前技术条件武器自主性带来的必然问题,适用可预见性标准意味着根本不可能存在金钱赔偿的问题。

综上所述,以不作为违反预防义务为路径,依旧不能完全解决超设定攻击违反国际人道法问题。2017年,美国在GGE会议的工作文件也持此立场。^⑩不可预见的超设定攻击导致的违法责任可能存在两个发展方向:其一,在未来的规则中确定严格责任路线,引入伦理风险、责任真空风险的可能性作为追求个人刑事责任、国家责任的基础;其二,军控机制未能对责任真空问题作出任何安排,归责难题属于次级规则领域而被军控谈判有意忽略。不过可以确定的两个趋势:一是全流程监管的落实与责任真空是负向影响关系,前者落实的越严密有效,不可预见的超设定攻击引发责任真空的概率越小;二是学理上引入严格责任有必要性,有助于封堵漏洞。

五、有意义的人类控制原则对我国的意义

2021年12月,我国外交部向联合国递交了《中国关于规范人工智能军事应用的立场文件》,该文件

提出我国在法律伦理和技术安全领域的主张,其与有意义的人类控制原则存在较多的契合之处:“法律伦理上,各国研发、部署和使用相关武器系统应遵循人类共同价值观,坚持以人为本,秉持‘智能向善’的原则,遵守国家或地区伦理道德准则。各国应确保新武器及其作战手段符合国际人道主义法和其他适用的国际法,努力减少附带伤亡、降低人员财产损失,避免相关武器系统的误用恶用,以及由此引发的滥杀滥伤。技术安全上,各国应不断提升人工智能技术的安全性、可靠性和可控性,增强对人工智能技术的安全评估和管控能力,确保有关武器系统永远处于人类控制之下,保障人类可随时中止其运行。人工智能数据的安全必须得到保证,应限制人工智能数据的军事化使用。”^⑪上述立场宣示涉及国际人道法、军控裁军规则、武器开发等诸多方面,洞见深刻,发人深省。然而,如何在法律层面落实,需要进一步丰富有意义的人类控制原则的含义。

从学理角度看,有意义的人类控制原则应在未来规制自主武器法律规则中处于中枢性地位。加强有意义的人类控制存在两类思路:一是增强国家对武器系统控制的确定性,表现在对自主武器系统自开发到使用评估的全流程监管,使自主武器的使用反映操作员、指挥官的意图,尽最大努力缩小武器系统自主性带来的不确定范围;二是完善责任机制,防止责任由人及国家传递至机器。前者面临军控和开发技术领域的问题,或者面临的问题是既有的国际刑法和国家责任规则在应对人工智能技术带来挑战时,不能完美地解决超设定攻击带来的问题。但可以预见的情况是,在军控领域设计和执行越严格的制度,国际刑法及国家责任法律面临超设定攻击带来“无人负责”困境的可能性就越小。

有意义的人类控制原则对中国自身法律制度的完善具有相当参考意义,对于我国在相关国际法规则的谈判中博弈,具有一定的指导意义,具体归纳如下:

第一,武器开发、生产、装备应当有具体规范的流程,涉及人工智能技术运用的,应该做出不同于一般自动化武器的针对性安排。我国现行法律中,仅有《刑法》第370条规定以故意或过失提供不合格武

器装备的构成犯罪,但自主武器可靠性问题能否被解释为合格不合格问题仍需考察,仅由刑法规范调整不足以落实有意义人类控制原则视角下全流程监管理念。例如,《武器装备质量管理条例》第13条规定:“武器装备论证单位应当对论证结果进行风险分析,提出降低或者控制风险的措施。武器装备研制总体方案应当优先选用成熟技术,对采用的新技术和关键技术,应当经过试验或者验证。”这一规定体现了有意义的人类控制原则思想基础,但当前我国仍未制定针对自主武器的开发、装备流程的法规或规章。

第二,实现全流程监管,需要在训练、使用方面加强对指挥官、操作员的训练,使其正确理解自主武器等操作规范。另外,在作战使用和毁伤效果评估方面强调交战规则,针对自主武器系统的特性进行调整,也是可以努力的方向。

第三,鉴于自主武器技术与武器系统的可分离性,应当考虑将与自主武器相关的人工智能技术列入军控法规的规制范围。2020年6月20日,第十三届全国人大第十九次会议表决通过了加入联合国《武器贸易条约》的决定,意味着符合该条约要求的自主武器及其弹药,可能纳入贸易监管的范围内,但可应用于武器系统的人工智能技术并不在该条约的管制范围之内。笔者认为,未来我国在修改国内法规,如《军品出口管理条例》之时,将可应用于武器系统的人工智能技术列入管制清单,尽管可能超出《武器贸易条约》要求,但此修改内容符合条约宗旨和我国作为人工智能技术大国的国家利益。

第四,武器自主化将带来战争形态的改变,因此对新武器和新作战方式方法、对第36条的审查具有重要地位,审查核心应该是评估自主武器是否满足遵守区分原则、比例原则、预先警告原则的要求。第36条审查本质上属于《第一附加议定书》缔约国自行决定的事务,若在CCW的多边框架下构建统一审查机制,并不符合我国利益。

在责任领域,有意义的人类控制原则面临的问题是,通过何种国际刑法和国家责任法的理论路径,实现个人及国家的责任不得向机器转移的共识。应对超设定攻击带来的挑战,国际刑法面临的主要问

题是,不可预见的超设定攻击无法为操作员、指挥官预见,此种情况下强行追究操作员、指挥官的责任,将违背主、客观要件结合归罪的一般方法。在国际责任层面,依据《国家责任条款草案》第4条、第8条及不作为违反预防义务原理归责,均存在障碍。立足于维护自身利益,围绕责任真空问题展开博弈,我国应当考虑的主要问题包括:第一,立足人工智能技术发展趋势,客观评估超设定攻击违反国际人道法且无法追责的典型场景及伦理风险,以及是否有必要建立我国强制责任分配机制(如严格责任);第二,客观评价在多边体制下建立强制责任分配机制被主要军事大国接受的可能性;第三,若在多边体制下建立强制责任分配机制,是否对自身造成严重不利,或对其他军事大国造成不对称的军事优势。

注释:

①截至目前,CCW框架下尚未形成对“致命性自主武器系统”的定义,甚至有的国家反对确定为是谈判进行下去临时采用的工作性定义。各国在其法规或政策中采用“致命性自主武器”“自主武器系统”“自主武器”等各种表述。本文中各种表述具有相同含义。

② "ICRC statement of 13 May 2014" available at <<http://www.icrc.org/eng/resources/documents/statement/2014/05-13-autonomous-weapons-statement.htm>>; "Report of the ICRC Expert Meeting on Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects", available at <<http://www.icrc.org/eng/assets/files/2014/expert-meeting-autonomous-weapons-icrc-report-2014-05-09.pdf>> (last visited on 19 May 2019).

③ US Department of Defense Directive 3000.09, November 11, 2012, Section 4(a).

④ US Department of Defense Directive 3000.09, November 11, 2012, Section 4(a)(1)(c).

⑤ Defense Innovation Board, AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense, Primary Document, p. 8.

⑥ 红十字国际委员会:《就〈特定常规武器公约〉向致命性自主武器系统政府专家组发表的声明》,红十字国际委员会网站,2021年8月3日,<<https://www.icrc.org/zh/document/autonomous-weapons-icrc-recommends-new-rules>> (访问日期:2022年1月20日)。

⑦ Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autono-

mous Weapons Systems, CCW/GGE. 1/2019/3, Annex 4, p. 14.

⑧ See United Nations Institute for Disarmament Research, *The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence*, pp. 2-4.

⑨ Prosecutor v. Gotovina et al., IT-06-90-A, Trial Judgments, ICTY, para. 1898.

⑩ Prosecutor v. Gotovina et al., IT-06-90-A, Appeals Judgments, para. 58.

⑪ ICRC, *Autonomy, Artificial Intelligence and Robotics: Technical Aspects of Human Control*, CCW/GGE. 1/2019/WP. 7, para. 4.

⑫ Ibid., paras. 10-11.

⑬ Additional Protocol I, Article 51(3)(adopted by 77 votes in favor, one against and 16 abstentions) (CDDH, Official Records, Vol. VI, CDDH/SR. 41, 26 May 1977, p. 16.); Additional Protocol II, Article 13(3)(adopted by consensus, CDDH, Official Records, Vol. VII, CDDH/SR. 52, 6 June 1977, p. 134.); Memorandum of Understanding on the Application of IHL between Croatia and the SFRY, para. 6, extracted from Jean-Marie Henckaerts & Louise Doswald-Beck, eds., *Customary International Humanitarian Law*, Vol. 2 (Cambridge University Press, 2005), p. 107.

⑭ 在红十字国际委员会的解释中,“平民直接参与敌对行动”包括三层次的构成要件:(1)该行为必须很可能对武装冲突一方的军事行动或军事能力造成不利影响,或者致使免受直接攻击之保护的人员死亡、受伤或物体毁损(损害下限);(2)在行为与可能因该行为(或该行为作为有机组成部分的协同军事行动)所造成的损害之间必须存在直接的因果关系(直接因果关系);(3)该行为必须是为了直接造成规定的损害下限,其目的是支持冲突一方并损害另一方(交战联系),参见厄尔斯·梅尔泽:《国际人道法中直接参加敌对行动定义的解释性指南》,国际红十字委员会官方出版物2009年版,第44页及以下。

⑮ Michael N. Schmitt, "Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics", (1) *Harvard National Security Journal Features* (2013), pp. 22 ff.

⑯ Michael N. Schmitt, "Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics", (1) *Harvard National Security Journal Features* (2013), pp. 19-20.

⑰ Report of the 2018 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, CCW/GGE. 1/2018/3, para. 21.

⑱ US Department of Defense Directive 3000. 09, November 11, 2012, Section 4(a).

⑲ Russian Federation, *Examination of Various Dimensions of Emerging Technologies in the area of Lethal Autonomous*

Weapons Systems, in the Context of the Objectives and Purposes of the Convention, CCW/GGE. 1/2017/WP. 8, paras. 12.

⑳ Position Paper of China, CCW/GGE. 1/2018/WP. 7, para. 3.

㉑ Japan, *Possible Outcome of 2019 Group of Governmental Experts and Future Actions of International Community on Lethal Autonomous Weapons Systems*, CCW/GGE. 1/2019/WP. 3, paras. 15-16.

㉒ Report of the 2018 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, CCW/GGE. 1/2018/3, Annex III, para. 10.

㉓ Report of the 2018 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, CCW/GGE. 1/2018/3, para. 21(b).

㉔ Report of the 2017 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS), CCW/GGE. 1/2017/CRP. 1, para. 16(c).

㉕ See Australia, *Australia's System of Control and Applications for Autonomous Weapon Systems*, CCW/GGE. 1/2019/WP. 2/Rev. 1, para. 16; Russian Federation, *Potential Opportunities and Limitations of Military Uses of Lethal Autonomous Weapons Systems* (English translate version), CCW/GGE. 1/2019/WP. 1, para. 8.

㉖ Japan, *Possible outcome of 2019 Group of Governmental Experts and Future Actions of International Community on Lethal Autonomous Weapons Systems*, CCW/GGE. 1/2019/WP. 3, para. 22.

㉗ Yves Sandoz et al., eds., *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (Martinus Nijhoff, 1987), p. 424.

㉘ Ibid., p. 428.

㉙ Ibid., p. 424.

㉚ Russian Federation, *Potential Opportunities and Limitations of Military Uses of Lethal Autonomous Weapons Systems* (English translate version), CCW/GGE. 1/2019/WP. 1, para. 8.

㉛ US Department of Defense Directive 3000. 09, November 11, 2012, Section 4(a)(3).

㉜ US Department of Defense Directive 3000. 09, November 11, 2012, Enclosure 2(a).

㉝ Australia's System of Control and Applications for Autonomous Weapon Systems, CCW/GGE. 1/2019/WP. 2/Rev. 1, paras. 30-31.

㉞ Ibid., para. 36.

㉟ US Department of Defense Directive 3000. 09, November 11, 2012, Section 4(d).

㊱ France, *Human-Machine Interaction in the Development,*

Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, CCW/GGE. 2/2018/WP. 3, para. 6; Christof Heyns, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, UN Doc. A/HRC/23/47, para. 79; 张卫华:《人工智能武器对国际人道法的新挑战》,《政法论坛》2019年第4期,第151-152页。

③⑦ Ndahimana Appeal Judgment, para. 38; Perišić Appeal Judgment, para. 87; Orić Appeal Judgment, para. 20; Čelebići Appeal Judgment, paras. 197-198.

③⑧ "Report of Commission on the Responsibility of the Authors of the War and on Enforcement of Penalties", 14(1/2)American Journal of International Law(1920), p. 141.

③⑨ Prosecutor v. Bemba Gombo, Pre-trial Chamber Decision on Confirmation of Charges, ICC-01/05-01/08, 15 June 2009, para. 425; Prosecutor v. Ntaganda. Decision on the Confirmation of Charges, Case No. ICC-01/04-02/06, 9 June 2014, para. 174; See Prosecutor v. Bemba Gombo, ICC-01/05-01/08, Trial Judgment, paras. 736-740.

④⑩ Prosecutor v. Zejnil Delalić et al., IT-96-21-T, paras. 437 and 439; Prosecutor v. Strugar, Appeal Judgment, para. 277; Kordić and Čerkez Appeal Judgment, paras. 89-91; Kordić and Čerkez Trial Judgment, para. 361; Hadžihasanović and Kubura Trial Judgment, para. 59; Krajišnik Trial Judgment, para. 782; Nalitić and Martinović Trial Judgment, paras. 603-605, citing Kordić and Čerkez Trial Judgment, para. 358 and Blaškić Trial Judgment, para. 185.

④⑪ Prosecutor v. Katanga, ICC-01/04-01/07, Trial Judgment, paras. 792, 793; Prosecutor v. Bemba, ICC-01/05-01/08, Trial Judgment, paras. 95-96.

④⑫ Prosecutor v. Al Mahdi, ICC-01/12-01/15, Judgment and Sentence, paras. 12-13.

④⑬ Prosecutor v. Katanga, ICC-01/04-01/07, Trial Judgment, paras. 898-890.

④⑭ 特别报告员克劳福德(James Crawford)提交给国际法委员会的草案文本中,同时考察了国际法院尼加拉瓜案(Nicaragua v. United States of America)和ICTY的塔迪奇案(Prosecutor v. Duško Tadić, 审判庭1998年判决,文本提交时正在上诉,审判庭判决采纳的标准与尼加拉瓜案保持一致),下文提及全面控制标准的提出是ICTY在1999年的上诉庭判决中明确提出(参见James Crawford, First Report on State Responsibility, UN Doc. A/CN.4/490/Add.5, pp. 17-20.)。因此,国际法委员会在2001年通过草案二读文本时认为,不论满足指令、指挥或控制哪个标准,均满足将非国家实体行为归因于国家的条件,有效控制标准和全面控制标准均可以作为归因性标准[参见ILC

Commentaries to the draft articles on Responsibility of States for internationally wrongful acts, extract from the Report of the International Law Commission on the work of its Fifty-third session, Official Records of the General Assembly, Fifty-sixth session, Supplement No. 10(A/56/10), chp. IV. E. 2, p. 108],换言之之根据草案文本理解,二者之间并不是非此即彼的关系。

④⑮ Case Concerning Military and Paramilitary Activities in and against Nicaragua, I. C. J. Judgment of Merits(1986), para. 115; Case Concerning the Application of Genocide Convention, I. C. J. Judgment(2007), para. 400.

④⑯ See Prosecutor v. Duško Tadić, Appeal Judgment, paras. 116-122.

④⑰ Tadić Appeals Judgment 15 July 1999, para. 141.

④⑱ Tadić Appeals Judgment 15 July 1999, para. 137; Aleksovski Appeals Judgment 24 March 2000, para. 147; Mucic et al.(“Čelebići”)Appeals Judgment 20 February 2001, para. 20.

④⑲ Case Concerning the Application of Genocide Convention, I. C. J. Judgment(2007), paras. 404-406.

④⑳ Prosecutor v. Jean-Pierre Bemba Gombo, Trial Judgment 21 March 2016, paras. 128-130; Katanga Trial Judgment 7 March 2014, para. 1178; Prosecutor v. Lubanga, No. ICC-01/04-01/06, Trial Judgment, para. 542.

⑤① Prosecutor v. Duško Tadić, Appeals Judgment, IT-94-1-A, paras. 126-127, citing Kenneth P. Yeager v. Islamic Republic of Iran, 17Iran-U. S. Claims Tribunal Reports, 1987, vol. IV, p. 92.

⑤② Tadić Appeals Judgment, IT-94-1-A, para. 128.

⑤③ Ibid., para. 129.

⑤④ Prosecutor v. Lubanga, Pre-trial Chamber Decision on the Confirmation of Charges, para. 211; Prosecutor v. Lubanga, Trial Judgment, ICC-01/04-01/06, 14 March 2012, para. 541.

⑤⑤ Prosecutor v. Katanga, Trial Judgment, ICC-01/04-01/07-3436, 7 March 2014, paras. 1178-1215.

⑤⑥ Encana Corporation v. Republic of Ecuador, London Court of International Arbitration Case No. UN3481, award, 3 February 2006, para. 154.

⑤⑦ WTO Appellate Body Report, United States—Countervailing Duty Investigation on Dynamic Random Access Memory Semiconductors(DRAMS)from Korea, WT/DS296/AB/R, 27 June 2005, para. 69.

⑤⑧ Československa Obchodní Banka A. S. v. The Slovak Republic, ICSID Case No. ARB/97/4, Decision on Objections to Jurisdiction, 24 May 1999(CL-053), para. 20.

⑤⑨ Emilio Agustín Maffezini v. The Kingdom of Spain, ICSID, Case No. ARB/97/7, Award, paras. 51-57.

⑥ Beijing Urban Construction Group Co. Ltd v. Republic of Yemen, ICSID Case No. ARB/14/30, Decision on Jurisdiction, paras. 31–36.

⑦ Saint-Gobain Prefrance Plastics Europe v. Bolivarian Republic of Venezuela, ICSID Case No. ARB/12/13, Decision on Liability and the Principle of Quantum, 30 December 2016, para. 453.

⑧ Ampal– American Israel Corporation and others v. Arab Republic of Egypt, ICSID Case No. ARB/12/11, Decision on Liability and Heads of Loss, 21 February 2017, para. 146.

⑨ Hulley Enterprises Limited(Cyprus) v. The Russian Federation, PCA Case No. AA 227, Final Award, 18 July, 2014, para. 1466.

⑩ Marfin Investment Group Holding S. A., Alexandros Bakatselos and Others v. Republic of Cyprus, ICSID Case No. ARB/13/27, Award, 26 July 2018, para. 691.

⑪ Case Concerning the Application of Genocide Convention, Judgment of 2007, I. C. J. Report 2007, paras. 163, 173, 182.

⑫ Ibid., paras. 209, 210.

⑬ The Corfu Channel Case(UK v. Albania), I. C. J. Reports of Judgments, Advisory Opinions and Orders(1949), p. 22.

⑭ Case Concerning Armed Activities on the Territory of the Congo(DRC v. Uganda), I. C. J. Judgment of 2005, paras. 247–250.

⑮ The Corfu Channel Case(UK v. Albania), I. C. J. Reports of Judgments, Advisory Opinions and Orders(1949), pp. 18–22.

⑯ Case Concerning Armed Activities on the Territory of the Congo(DRC v. Uganda), Judgment, paras. 174–180.

⑰ Case Concerning the Application of Genocide Convention, Judgment of 2007, para. 431.

⑱ Ibid.

⑲ Ibid.

⑳ James Crawford, State Responsibility: The General Part (Cambridge University Press, 2013), p. 221.

㉑ Ibid.

㉒ Case Concerning the Application of Genocide Convention, Judgment of 2007, para. 430.

㉓ Lassa Oppenheim, International law: a treatise(Longmans, Green & Co., 1920), pp. 258–259.

㉔ Scott J. Shackelford & Richard B. Andres, "State Responsi-

bility for Cyber Attacks: Competing Standards for A Growing Problem", 42(4)Georgetown Journal of International Law(2011), pp. 989–990.

㉕ Case Concerning Diplomatic and Consular Staff of United States(US v. Iran), Judgment of 1979, I. C. J. Report 1980, para. 74, p. 35.

㉖ Case Concerning Diplomatic and Consular Staff of United States(US v. Iran), Judgment of 1979, I. C. J. Report 1980, paras. 63–68, pp. 31–33.

㉗ ILC, Commentaries to the Draft Articles on Responsibility of States for Internationally Wrongful Acts, [extract from the Report of the International Law Commission on the work of its Fifty-third session, Official Records of the General Assembly, Fifty-sixth session, Supplement No. 10(A/56/10), chp. IV. E. 2], p. 120.

㉘ ILC Draft Articles on Responsibility of States for Internationally Wrongful Acts(first reading)with Commentaries, p. 174.

㉙ Roberto Ago, Second Report on State Responsibility, A/ CN/233, extract from the Yearbook of the International Law Commission(1970), Vol. II, p. 194.

㉚ See ILC Draft Articles on Responsibility of States for Internationally Wrongful Acts(first reading)with Commentaries, pp. 173–177.

㉛ James Crawford, State Responsibility: The General Part (Cambridge University Press, 2013), p. 493.

㉜ 参见郑斌:《国际法院与法庭适用的一般法律原则》,韩秀丽、蔡从燕译,法律出版社2012年版,第260页。

㉝ Application of Genocide Convention, para. 462.

㉞ Application of Genocide Convention, para. 462.

㉟ Prosecutor v. Germain Katanga, Decision on the Matter of the Transgenerational Harm Alleged by Some Applicants for Reparations Remanded by the Appeals Chamber in its Judgment of 8 March 2018, ICC–01/04–01/07, 19 July 2018, para. 17; Eritrea–Ethiopia Claims Commission(Ethiopia v. Eritrea)Final Award, Ethiopia's Damages Claims, 17 August 2009, para. 292.

㊱ USA, Autonomy in Weapon Systems, CCW/GGE. 1/2017/WP. 6, para. 30.

㊲ 参见《中国关于规范人工智能军事应用的立场文件》,外交部军控司网站,2021年12月14日, <http://infogate.fmprc.gov.cn/web/wjlb_673085/zzjg_673183/jks_674633/jksxwlb_674635/202112/t20211214_10469511.shtml>(访问日期:2021年12月30日)。