人工智能应用于军事的伦理问题

何杰[1],孙啸宇[2],郑睿[3],王兰英[1*]

1. 大连医科大学学生处,大连 116044
2. 大连医科大学人文与社会科学学院,大连 116044
3. 大连医科大学附属第一医院,大连 116044

摘要 近年来,人工智能军事化进程正随着复杂多变的国际局势高速发展。从人工智能的概念及其伦理风险出发,列举了国内外军用人工智能发展中具有代表性的伦理问题,包括身份识别技术与道德判断、责任归属、由人工智能失控引发的灾难问题等。分析了人工智能军事化伦理风险产生的根本原因,提出了以确保避免军用人工智能造成人类重大灾难为底线、持续改善军用人工智能存在的问责差距与技术漏洞、努力推进国际之间军用人工智能话题的积极讨论等3方面建议。

关键词 人工智能;军事应用;伦理困境;伦理审视

自从1955年McCarthy等[1]在达特茅斯首次提出人工智能这一概念以来,经过半个多世纪的发展,在算力、大数据和深度学习等技术发展的推动下,人工智能取得了显著进步,彻底改变了图像识别、自然语言处理和语音合成等应用方向。在民用领域,人工智能通过语音助手、自动驾驶、推荐系统、搜索引擎和神经网络翻译等深度介入日常生活。恩格斯曾指出"一旦技术上的进步可以用于军事目的并且已经用于军事目的,它们便立刻几乎强制地,而且往往是违反指挥官的意志而引起作战方式上的改变甚至变革"[2]。

近年来,具有一定智能的武器系统已经投入实战,数量、质量、规模迅速增长,战法、战术、任务内容日益丰富。2015年12月,世界上第1场以战斗机器人为主的攻坚作战诞生于叙利亚战场,俄罗斯将自主机器人部队投入拉塔基亚754.5高地参与作战[3]。2020年3月,在利比亚内战中,政府军一架土耳其"卡古-2"无人机使用自主攻击模式杀死人

收稿日期:2023-05-22;修回日期:2023-09-18

基金项目:中国指挥与控制学会国防教育研究专项一般项目(GFJY2022YB007);辽宁省教育厅基本科研项目(JYTMS20230593);大连市社科联立项课题(2022dlskzd283);度大连医科大学教学改革项目(DYLX23025);中央军委军事理论重点项目

作者简介:何杰,讲师,研究方向为军事学、国防教育,电子信箱:hj0326@163.com;孙啸宇(共同第一作者),硕士研究生,研究方向为医学伦理、心理学,电子信箱:2578816057@qq.com;王兰英(通信作者),副教授,研究方向为医学伦理、国防教育,电子信箱:709998691@qq.com

引用格式:何杰,孙啸宇,郑睿, 等. 人工智能应用于军事的伦理问题[J]. 科技导报, 2024, 42(4): 124-132; doi: 10.3981/j.issn.1000-7857.2024.04.013

类目标[4]。在最近的俄乌冲突中,乌克兰首次将面部识别技术用于物理进攻、信息作战、身份识别等[5]。随着人工智能在军事行动中自主性的逐渐增加,其伦理风险越来越不容忽视。

1　人工智能的概念及其伦理风险的体现

McCarthy等[1]将人工智能定义为“制造智能机器的科学和工程”。这一定义强调:以创造能够表现出智能行为,并执行通常需要人类智能的任务的机器为目标。McCarthy的定义为人工智能领域奠定了基础,并对其发展和研究产生了深刻影响。

以目前的技术水平,创建人工智能系统通常有一系列步骤,包括明确需求、收集及预处理数据训练模型、部署及运维。从创建过程看,人工智能的伦理风险主要体现在数据偏见和算法误差。人工智能系统可以从训练的数据中继承偏见,这可能导致歧视性结果,例如,针对特定群体或在军事行动中延续现有偏见。一个名为Tay的聊天机器人,通过推特和其他社交平台模仿,学习一个19岁问题青年的言论,该机器人在短短1 h内发表带有种族色彩的不当言论,宣扬纳粹意识形态,并开始骚扰其他用户[6]。这一问题目前看难以通过技术进步解决。算法误差表现为现有模型的可解释性差,这就导致其难以被人类理解并可能表现出不可预测的行为。现有人工智能系统对人类交战规则的“理解”和“感知”与人类不同,可能在执行过程中出现错误、故障或被对手利用,这可能导致重大的道德困境和潜在的伤害。但这一问题可能随着技术进步而有所改善。

2　人工智能军事化的伦理困境

英国物理学家霍金曾表示出对人工智能未来发展的担忧,认为人工智能的广泛使用具有极高的风险性,它的“超人类性”和“非可控性”可能促使人类文明走向终结[7]。导致这种结果的首要原因是,人工智能当前与今后在军事中的发展缺乏规范。目前人工智能在军事领域的应用主要有:自主武器系统、网络安全、指挥控制系统、物流和供应链管理、目标识别、自然语言处理、决策支持系统、训练和模拟等方面。

2.1　身份识别技术与道德判断

目前人们对于人工智能的探索仍处于早期阶段,存在技术的局限性与两面性,大量伦理问题由此产生。当军用人工智能设备基本或完全脱离人类操控,无法准确识别战斗人员与平民是技术局限性的重点问题。作为战争正义性的基本原则,区分原则要求严格区分军事人员与平民、军用设施与民用设施,避免对平民及其生存环境造成不必要的破坏。然而,目前人工智能的识别技术会出现将河马识别成公交车、将拐杖识别为步枪等错误,使用人工智能武器系统的最大阻碍在于,无法可靠地识别事物的合法性[8]。士兵可能通过经验或感觉锁定一个伪装成平民的恐怖分子,但人工智能无法做到。技术支持者提出,自主武器系统通过编程,将逐渐获得关于人类如何解释和预测行为的逻辑,例如,“通过使用步态识别和其他活动模式来识别可疑人员”[9],以及通过结合使用射频识别(RFID)之类的技术以帮助区分敌友。此后,Lin等[10]提出了更广泛的内容,诸如面部表情、凝视方向、肢体语言、服装、主体在环境中运动的信息、主体感觉器官的信息、主体的背景信念、欲望、希望、恐惧和其他精神状态的信息等。事实上,真实的战场环境远比想象的更复杂,当机器人看到2个小孩拿着刀朝着它们跑来,它们会基于什么算法认为孩子存在威胁?此外,敌人将利用人工智能的识别局限性,更倾向于将军队聚集在民用设施和平民当中,而用于识别身份的RFID标签会面临出现故障和被克隆等风险[11]。但另一方面,无法要求每一名士兵都能够在紧张的战场环境中做出准确的判断,在伊拉克战场中,近7%的士兵在没有任何理由的情况下伤害了非战斗人员[12]。这里不仅涉及士兵战场形势判断能力的问题,同时也反映出士兵道德判断能力的问题。

战争正义性的另一重要原则为比例原则,即根据作战目的和实际情况使用相称的暴力手段,以避

免造成过度的破坏和伤亡。在实际作战环境中,战斗人员对生命价值和具体形势的主观判断决定了军事行动是否符合“比例原则”的道德评估,是一种高度复杂、难以复刻的思维过程。Arkin等[9]认为,通过对系统的感知和行为进行编码,最终能够创造出控制道德行为的框架。将道德能力融入机器人系统可能包括道德词汇、规范体系、道德认知与情感、道德决策与行动、道德沟通[13]。然而,人类对世界的感知是整体的,且道德规范本身十分复杂,并且会随着时间的推移、空间的转换而变化,难以实现普世性[14]。更关键的是,人工智能无法理解人类生命的价值,缺乏感知自己或他人死亡的能力[15]。综上,权衡好军用人工智能技术在现实战场环境中的优势和劣势,是人工智能在未来战场中投入使用的重要前提。

2.2 责任归属问题

在特定战场环境中,虽然人类士兵作为道德主体可能会面临道德困境并做出违背道德的行为[16],但并不意味着应当接受人工智能系统在面临道德困境时做出的所有决策。因为人工智能系统不能作为道德主体,无法让人工智能系统对自己的行为负责[17]。如果工人习惯性迟到,并且没有被追究责任,那么它们很可能会继续迟到,其他人很可能也会跟进,最终准时上班将不再是常态。在国际上,人类拥有战争的权利,必须以承担战争责任为前提,这关乎到民族尊严和国家荣誉[18]。一名士兵在战场上犯错,就要为其后果负责。然而,当未来战争以人工智能为主导,战争责任由谁承担、如何承担是目前需要国际间达成共识的重要议题。如果人工智能没有明确的问责方案,甚至免除人类对某些人工智能违规行为的责任,就会鼓励人们更有恃无恐地使用人工智能,从而导致战争失控。

谷歌公司3000多名员工曾联名致信谷歌CEO(首席执行官),要求立即终止参与军用人工智能相关项目,声明永不研发战争技术[19]。很多微软员工也曾联名抗议公司与美国陆军签署合同[20]。这些事件并不会阻止美国发展军用人工智能的脚步,但随着因技术缺陷造成安全事故的增多,问责差距、问责真空的问题日益凸显。Wallach等[21]在《道德机器,如何让机器人明辨是非》中指出“随着环境变得更加复杂,且计算系统的内部处理需要管理大量变量,构建系统的设计师和工程师可能无法预测系统将遇到的许多情况或处理新信息的方式。”如果不可能从人类的决策中充分解释机器行为,那么就有可能出现无人负责的道德违规行为。因此,无法充分解释机器行为会导致责任差距,破坏战争惯例,并使战争失去人性[22]。对此,美国不得不推进军用人工智能的问责机制。

设计师和工程师是军用人工智能投入使用的第一环节人员。Whetham[23]认为,行动的合法性取决于正确的动机。意图是道德责任的必要条件,如果没有意图,就不可能有责任,那么自主武器系统的设计必须包含动机[24]。然而,自主武器系统的行为在多大程度上能够代表研发者的动机和意图,却受制于技术的局限性。Chmielewski[25]曾在技术的研发层面给出建议,第一,在设计和生产层面,一系列利益相关者将参与其中;第二,必须允许操作员和调度员参与有关预期和投入使用后存在问题的批判性讨论;第三,必须设计智能网络元素,以便参与自身的道德评估。同时,参与设计的团队必须包括来自人类认知学科和社会学科的专家作为系统开发的关键成员。此外,设计必须考虑地理、历史和文化差异的判断变量。因此,系统设计必须利用机器学习来掌握覆盖目标坐标的历史和文化因素,特别是识别这些因素与使用自主武器的社会因素的差异。上述建议较为完整地规范了研发军用人工智能过程中的要求,但人类现有的能力似乎无法完成如此庞大、复杂的工程。

指挥官和操作员是军用人工智能投入使用的中间环节和最后一环人员。指挥官的责任分为2方面,一方面,他们要对下达的命令负指挥责任,意味着无论他们是否有意,都要对本应避免的错误命令负责。正如《纽伦堡法典》所述,“如果指挥官未能要求并获得完整的信息,那就是他的失职,且无法为自己的失职辩护。”另一方面,他们需要了解其下属的行为。因此,设计师必须尽最大努力确保人工智能系统的输出对指挥官和操作员是“可解释的”[26]。虽然,这种解释的程度可能无法实现完全

绝对理解,尽管没有一个指挥官完全知道自己的下属在做什么,他们仍然能够对下属犯下的任何违法行为负责。因此,对指挥官来说,重要的是他们是否信任这些下属在可能违反战争惯例的情况下能够做出合乎道德的行为。对于军用人工智能,从责任合理分配的角度,首先,采购官员、设计师、程序员、制造商,以及指挥官和操作员必须牢记战争惯例,并履行其职责;其次,指挥官和操作员不仅必须了解设备在做什么,而且必须充分了解设备的工作方式,以便更好地理解设备如何解释、执行指令并提供输出;再次,在可能的范围内,指挥官和操作员必须能够监控设备的运行并防止其发生错误,从而防止机器违规;最后,操作员无法干预的系统应至少能够让指挥官和操作员相信他们能像人类士兵一样执行任务[27]。因此,如果指挥官相信人工智能系统能像人类士兵一样在道德选择方面发挥作用,那么使用这些系统才能在道德上被允许。

2.3 由人工智能失控引发的灾难问题

"有意义的人类控制"是为应对自主性武器安全问题而提出的理念[28]。大部分国家和组织均认可该理念是国际社会就自主性武器军控达成共识的潜在基础[29-30]。国际机器人武器控制委员会(International Committee for Robot Arms Control, ICRAC)等认为,"有意义的人类控制"必须要求人类操作员对目标区域有充分的语境和情境意识。他们还需要足够的时间来考虑目标的性质、攻击的必要性和适当性,以及可能的附带危害和影响。如果需要满足其他条件,他们必须具备中止攻击的手段[31]。"有意义的人类控制"等级体现在人类与人工智能系统交互的3种方式上。按照控制程度从强到弱,分别为人在回路中、人在回路上和人在回路外[32]。这里的"回路"是指设备相对于特定目的执行的"感知决定行为"操作。当人在回路中,机器在执行任务时等待人的输入操作。当人在回路上,机器可以自己感知、决定和行动,但人类可以监控系统并随时进行干预以防止其做出意料之外的行动。当人在回路外,机器在没有人类监督的情况下自行感知、决定和行动。事实上,当人在回路外时就已经脱离了有意义的人类控制,在上述责任归属问题尚未解决之前,人在回路外的设计方案并不可行。而当人在回路中时,相当于操作员手握操纵设备的遥控器,这会增加操作员杀人的敏感性,从而增加创伤后应激障碍、精神伤害等可能性。2015年,美国大量无人机操作员辞职,其中一些人源于过度劳累,另一些人源于他们认为应该为恐怖事件负责[33]。相比于传统战场,虽然与战场的距离缓解了士兵对战争的恐惧,但人在回路中引起的杀人责任带来的情绪问题依然需要尽可能缓解。当人在回路上,操作员只对设备实施监控,并在必要的时候进行干预。这似乎解决了操作员所面临的情绪问题,但其本身仍然存在新的问题。人在回路上的问题是,即使人类可以防止错误的机器行为,但他们通常不会这样做。这种反直觉的结果源于人类将设备所呈现的对事物的判断视为事实。总之,目前尚无法通过选择上述3种控制方式中的某一种来规避全部问题。然而,从全局考虑,人工智能脱离人类控制的程度越高,人类所面临的风险也会随之增大。

近年来,美国已多次将军用人工智能应用于反恐行动中。军队与恐怖分子之间的战斗很快将转变为更多的遥控武器装备和更少的人员参与,这一定程度上提升了行动效率,缓解了人员伤亡问题,但恐怖组织利用人工智能的反扑似乎也将接踵而至。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"?

结合国际公约、中国法律,以及刑法学界对恐怖主义的界定,恐怖主义的本质是为了实现特定目的与动机(大多数为了追求精神上的满足),通过暴力或非暴力的手段,制造恐怖氛围、社会恐慌,干扰政府行动,严重威胁相关人员的财产与人身安全[34]。恐怖分子犯罪动机和结果的特殊性如果与人工智能技术的颠覆性相结合,必将对现在的人类社会带来威胁。人工智能与核武器或其他传统武器不同,并不需要昂贵的、难以获取的原材料,以及受到严格管控的绝密技术。可以设想未来恐怖分子使用人工智能实施犯罪大体分为3种情况:其一,恐怖分子利用人工智能实施爆炸、暗杀、劫持人

质等传统恐怖主义犯罪;其二,将人工智能与网络平台结合,实施煽动性、传播性恐怖主义活动;其三,使用智能机器人(人在回路外)实施恐怖主义活动,试图危害甚至毁灭人类。虽然上述3种情况目前暂未发生于现实世界里,但这种貌似科幻的设想很有可能在未来的某天变成现实。恐怖分子可能在未来轻易掌握军民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,进而增加对国际安全构成的潜在风险[18]。此外,在计算机科学与技术和3D打印等高新技术的推动下,获得杀手机器人将更加轻而易举[35]。至少目前看,人工智能系统的思维是一个"黑匣子",人类没有能力确保将这项技术牢牢掌控。没有什么能阻止黑市的人设计军用机器人,并将其"坏"的价值体系灌输于处于道德规范阶段的机器。同样,当一个人可以教机器人某一特定行为是坏的,另一个人当然可以教机器人这样的行为是好的[36]。如果人工智能武器的设计者怀有种族灭绝、恐怖主义、反人类等犯罪心理,人类将无法承担其带来的灾难后果。

3 人工智能军事化伦理风险产生的根本原因

3.1 自我意识存在的进退两难

人工智能的本质是算法、模型和参数,无法真正理解或感知。目前基于深度学习的人工智能本质是利用大量数据在高维数学空间中进行层层神经网络运算的计算机程序,人工智能模型通常被视为函数近似,而不是像人类思维过程那样可以直接理解或感知。

通常认为当前的人工智能系统不存在自我意识。在缺乏自我意识的情况下,其中一个重要的问题是偏见和歧视的持续存在。人工智能系统即使没有自我意识,也可能会无意中继承和强化存在于社会中的偏见。此外,人工智能缺乏主观意识,带来了与问责制和责任相关的挑战。目前人工智能系统可能难以解释其决策过程,因此阻碍了追究人工智能武器错误的或有偏见的行为结果的责任。

虽然人工智能存在自我意识这一观点仍然只是推测,但可以预见,终有一天人工智能会发展出自我意识。自我意识为人工智能武器系统行为的不可预测性带来了重大风险。其独立思考能力引入了不可预测性的因素,可能导致违背人类价值观和意图。如果具有自我意识的人工智能与人类之间的利益不一致,会进一步加剧这些担忧。如果具有自我意识的人工智能能够超越人类智能,人们就会担心它们的行为可能会对人类的生存构成威胁。这些担忧共同强调了识别和减轻人工智能自我意识的伦理影响至关重要。

3.2 可解释性差

许多高级人工智能模型,特别是深度学习模型,由无数层和数百万或数十亿个参数组成。这些模型中的相互作用和计算非常复杂,人类很难解释。人工智能模型通常捕捉输入特征和输出结果之间的非线性关系。人工智能模型通常在高维特征空间中运行,对人类而言很难可视化或概念化,也使理解不同的特征如何对模型的预测产生影响变得很有挑战性。深度学习模型通常被视为"黑匣子",因为内部工作不容易解释。虽然可以观察输入和输出,但模型中发生的具体计算和转换很难辨别。人工智能模型从大量数据中学习,并提取对人类来说可能并不明显的模式,所获得的见解是基于统计相关性,而不是明确的基于规则的逻辑,没有办法进行直观的解释。在许多情况下,人工智能模型会自动从原始数据中学习表现和特征,这些习得的表征可能与人类可理解的概念不一致,因此很难提供明确的解释。

一些人工智能模型,特别是在复杂环境和高度不确定性环境中使用的模型(如军事行动中),本质上存在不确定性。这使得为它们的结果提供明确的解释变得很有挑战性。一些人工智能系统具有适应和持续学习的能力,这意味着它们的行为可以随着时间的推移而进化。这种动态性会使人们很难预测或解释它们在未来所有可能的场景中的行为。人工智能模型还可能会无意中学习训练数据中存在的偏见,识别和解释这些偏见同样具有挑战性,尤其当它们是微妙的或依赖于语境的。

3.3 人工智能无法作为责任主体

人工智能从根本上缺乏真正负责所需的基本属性。与人类不同,人工智能缺乏主观意识、情感和意图。它在预定义的算法和数据模式的范围内运行,缺乏真正的道德。人工智能的确定性本质意味着其行为完全由编程决定,不具备自主道德推理的能力。此外,人工智能缺乏在道德或伦理背景下理解其行为的更广泛影响或后果的能力。它不能拥有个人身份或责任,而这些是真正道德责任的关键因素。虽然人工智能精通基于数据驱动算法的任务,但它无法真正地进行道德判断或伦理思考。因此,人工智能行为的最终责任是作为其创造者和操作者的人类,他们需要对人工智能系统的设计、实施和结果负责。总之,人工智能的固有属性和能力使其无法承担道德和伦理意义上的真正责任。

4 人工智能军事化的伦理审视及未来发展

军用人工智能投入战场的诸多现实案例正在警醒人类对这项技术的考量,同时也反映出智能化战争的脚步已经迫近,人工智能的军事化发展趋势不可逆转。当前的国际形势,美国借助军事技术发展的领先地位,通过积极表达立场、推行美国经验、阐释“美式理论”等方式,引导人工智能武器国际讨论的走向,努力确保美国在该问题上的话语权和领导力[37]。对此,我们不得不积极应对军用人工智能发展过程中的种种挑战,结合实际需要,提出更科学、有效的相关伦理原则势在必行。

4.1 以确保避免军用人工智能造成人类重大灾难为底线

从全局角度出发,一项技术想要科学稳定地发展,必须始终确保其不会存在毁灭性破坏的可能。回顾军用人工智能发展的诸多伦理问题,不难发现,最有可能造成灾难性后果的是来自恐怖主义的威胁,而这种威胁的核心在于我们能否牢牢把握军用人工智能的控制权。虽然当前的技术水平距离“人在回路外”的人机交互方式仍然很远,介于无法界定责任归属、难以赋予道德判断,最重要的是无法对人工智能武器进行有效控制。因此,应当明确,在人工智能的“黑匣子”没有被完全打开之前,“人在回路外”的人工智能武器的研发和应用应该予以禁止。同时,要切实有效地规范“人在回路中”“人在回路上”的人机交互模式,把人类对人工智能的安全控制放在制定、研发、投入使用过程中的首要位置。

4.2 持续改善军用人工智能存在的问责差距、技术漏洞

人工智能不同于其他前沿科技,其极高的安全风险贯穿于产品的制定、研发、生产和使用阶段,因此,每一阶段的行事准则都要严格规范。严格选拔人才,组建符合人类文明发展需要的核心研发团队。完善监察机制,在确保技术研制安全、合规的基础上,尽可能促进团队在核心技术方面取得突破性成果。通过开展还原真实战场环境的无人化实战演习,落实、落细指挥员对各类人工智能的管理权限及操作员对所属人工智能的操作规范,逐步加强指挥员、操作员对未来战场中所属设备的信任。同时参照传统军事行动现行的责任制度,结合已研发军用人工智能的运行特征,制定符合军用人工智能实际的伦理道德专项制度,抵消问责差距。综上,结合当前国际局势及国内综合发展现状,应当有原则、有底线、积极地发展军用人工智能技术。

4.3 努力推进国际之间军用人工智能话题的积极讨论

2020年2月,美国国防部发布了人工智能军事应用的5项原则,表面上规范了自身军用人工智能的发展,究其本质,是为了在世界范围内抢占制高点做准备。其最终目的是以一种符合美国价值理念的方式,扫清舆论障碍、掌控国际军用人工智能讨论的话语权[38]。关于国际对自主武器安全问题的讨论,美国曾质疑国际普遍接受的“有意义的人类控制”理念,反之推崇“适当的人类判断”。此后,当发现无法改变国际共识,美国开始用有利于自身的价值体系解释“有意义的人类控制”理念,无不反映其单边思维和霸权逻辑[38]。对此,中国应当积极引导国际间的讨论,深度挖掘“有意义的人类控制”理念的理论基础和可操作性,广泛参与制定符合人

类共识的伦理准则。

5 结论

就像历史上发生过多次的工业革命,人工智能也是一次“工业革命”。当今人们对于人工智能担忧的本质,也像历史上发生过多次的那样,资本主义制度异化了劳动,进而异化了科学技术,从而造成了机器对工人的统治,以及工人对机器的抗争[39]。但是每一次生产力的发展都会带来生产关系的改变,人工智能技术作为一次“工业革命”带来了生产力飞跃,也带来了人类对未来生存发展安全的恐慌,需要以伦理的规范加以约束。从发展的眼光看,人工智能伦理与规范是未来智能社会的发展基石,算法偏见、安全性、可靠性、责任缺位等问题的凸显使制定人工智能伦理准则的必要性日益凸显。

同时,关于人工智能应用军事领域的伦理审视更加值得各国广泛关注。但需要警惕的是,美国国防部率先推出军事领域的人工智能伦理准则,既有规范美国军事人工智能发展与应用的目的,也是为了破除阻碍美国军事运用人工智能的舆论障碍,管控人工智能军事应用可能带来的内外部风险,同时为抢占军事人工智能伦理准则制定的国际主导权,维持美国科技和军事优势、赢得未来战争提早进行战略准备[38]。因此,需要站在批判的角度,科学分析美国推出的军事人工智能伦理准则,不断推进中国人工智能军事应用安全发展、加快军事智能化建设,积极参与人工智能国际安全治理。

参考文献(References)

[1] McCarthy J, Minsky M, Rochester N, et al. A proposal for the Dartmouth summer research project on artificial intelligence[J]. AI Magazine, 1955, 27(4): 12.

[2] 马克思, 恩格斯. 马克思恩格斯选集(第3卷)[M]. 北京: 人民出版社, 1972: 211.

[3] 马建光. 叙利亚战争启示录[M]. 武汉: 长江文艺出版社, 2021: 43.

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

[5] Dave P, Dastin J, Exclusive-Ukraine has started using Clearview AI's facial recognition during war[EB/OL]. (2022-03-13)[2023-11-22]. https://www.marketscreener.com/news/latest/Exclusive-Ukraine-has-started-using-Clearview-AI-s-facial-recognition-during-war-39750732/.

[6] Staff A. Tay, the neo-Nazi millennial chat bot, gets autopsied[EB/OL]. (2016-03-26)[2023-11-22]. https://arstechnica.com/information-technology/2016/03/tay-the-neo-nazi-millennial-chatbot-gets-autopsied.

[7] 黄志澄. 如何看待霍金对人工智能的警告[EB/OL]. (2017-06-05)[2023-11-22]. http://opinion.people.com.cn/n1/2017/0605/c1003-29316746.html.

[8] Umbrello S, Torres P, Bellis D. The future of war: Could lethal autonomous weapons make conflict more ethical[J]. AI & Society, 2020, 35: 273-282.

[9] Arkin R. Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture part I: Motivation and philosophy[C]//3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI). Amsterdam, Netherlands: IEEE, 2008: 121-128.

[10] Lin P, Abney K, Bekey G. Robot ethics: The ethical and social implications of robotics[M]. Cambridge: MIT Press, 2012: 129-144.

[11] Garfinkel L, Juels A, Pappu R. RFID privacy: An overview of problems and proposed solutions[J]. IEEE Security and Privacy Magazine, 2005, 3(3): 34-43.

[12] Office of the Surgeon Multinational Force-Iraq, Office of the Surgeon General (Army). Mental health advisory team (MHAT) IV operation Iraqi freedom 05-07. final report[R]. Washington DC: Army, 2006.

[13] Malle F. Integrating robot ethics and machine morality: The study and design of moral competence in robots[J]. Ethics and Information Technology, 2015, 18(4): 243-256.

[14] Christen M, Burri T, Chapa J, et al. An evaluation schema for the ethical use of autonomous robotic systems in security applications[R/OL]. (2017-10-12) [2023-11-22]. https://www.researchgate.net/publication/321091297_An_Evaluation_Schema_for_the_Ethical_Use_of_Autonomous_Robotic_Systems_in_Security_Applications.

[15] Skerker M, Purves D, Jenkins R. Autonomous weapons systems and the moral equality of combatants[J]. Ethics and Information Technology, 2020, 22(3): 197-209.

[16] Bellaby R. Can AI weapons make ethical decisions?[J]. Criminal Justice Ethics, 2021, 40(2): 86-107.

[17] 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(外交学院学报), 2018, 35(1): 128-156.
[18] 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(外交学院学报), 2022, 39(3): 99-130.
[19] Harwell D. Google to drop Pentagon AI contract after employee objections to the 'usiness of war'[EB/OL]. (2018-06-02) [2023-11-22]. https://www. ndtv. com/world-news/google-to-drop-pentagon-ai-contract-after-employees-called-it-the-business-of-war-1861298.
[20] Carbone C. Microsoft employees slam $480M HoloLens military contract, refuse to create tech for "warfare and oppression"[EB/OL]. (2019-02-24)[2023-11-22]. https://www.foxnews.com/tech/microsoft-employees-slam-480m-hololens-military-contract-demand-its-cancellation.
[21] Wallach W, Allen C. Framing robot arms control[J]. Ethics and Information Technology, 2013, 15(2): 125-135.
[22] Roff H. Killing in war: Responsibility, liability, and lethal autonomous robots[M]//Roff H. Routledge Handbook of Ethics and War. London: Routledge and CRC Press, 2013: 13.
[23] Whetham D. Morality and war: Can war be just in the twenty-first century?[J]. Scientia Militaria South African Journal of Military Studies, 2012, 40(1): 881-883.
[24] Sparrow R. Killer robots[J]. Journal of Applied Philosophy, 2007, 24(1): 62-77.
[25] Chmielewski P. Ethical autonomous weapons? Practical, required functions[J]. IEEE Technology and Society Magazine, 2018, 37(3): 48-55.
[26] Scharre C, Horowitz P. Artificial intelligence: What every policymaker needs to know[EB/OL]. (2018-06-01) [2023-11-22]. https://www.jstor.org/stable/resrep20447.1.
[27] Pfaff C A. The ethics of acquiring disruptive technologies: Artificial intelligence, autonomous weapons, and decision support systems[EB/OL]. (2020-01-10) [2023-11-22]. https://ndupress.ndu.edu/Media/News/News-Article-View/Article/2054156/the-ethics-of-acquiring-dis - ruptive-technologies-artificial-intelligence-autonom/.
[28] Article 36. Killer Robots: UK government policy on fully autonomous weapons[EB/OL]. (2013-04-21) [2023-11-22]. https://article36. org/wp-content/uploads/2013/04/Policy_Paper1.pdf1.
[29] 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[EB/OL]. (2021-01-09) [2023-11-22]. https://docs-library. unoda. org/Convention_on_Certain_Conventional_Weapons_ - _Group_of_Governmental_Experts_(2017)/2017_GGEonLAWS_WP6_USA.pdf.
[30] International Committee of the Red Cross. Views of the International Committee of the Red Cross on autonomous weapon system[C]//CCW Meetings of Experts on LAWS. Geneva: Group of Governmental Experts on Lethal Autonomous Weapons Systems, 2016: 3-4.
[31] Sharkey N. Guidelines for the human control of weapon systems[R]. Sheffield UK: International Committee for Robot Arms Control, 2018.
[32] Caton J L. Autonomous weapon systems: A brief survey of developmental, operational, legal, and ethical issues [R]. Carlisle USA: Strategic Studies Institute, US Army War College, 2015.
[33] Chatterjee P. American drone operators are quitting in record numbers[EB/OL]. (2015-03-15) [2023-11-22]. https://www.thenation.com/article/archive/american-drone-operators-are-quitting-record-numbers/.
[34] 王燕玲, 李瑞华. 人工智能时代恐怖主义犯罪行为的刑法规制[J]. 刑法论丛, 2020, 64(4): 134-160.
[35] Haner J, Garcia D, Held D, et al. The artificial intelligence arms race: Trends and world leaders in autonomous weapons development[J]. Global Policy, 2019, 10 (3): 331-337.
[36] Brown-Gaston R D, Arora A S. War and peace: Ethical challenges and risks in military robotics[J]. International Journal of Intelligent Information Technologies, 2021, 17 (3): 12.
[37] 王玫黎, 杜陈洁. 美国参与自主性武器国际军控的战略关切及角色定位[J]. 国际观察, 2021(2): 127-156.
[38] 龙坤, 徐能武. 对美军推出人工智能伦理准则的剖析[J]. 情报杂志, 2022, 41(3): 1-8.
[39] 李琼琼, 李振. 智能时代"人机关系"辩证——马克思"人与机器"思想的当代回响[J]. 毛泽东邓小平理论研究, 2021(1): 71-79, 108.

Ethical thinking on the application of AI in military

HE Jie[1], SUN Xiaoyu[2], ZHENG Rui[3], WANG Lanying[1*]

1. Students′ Affairs Office, Dalian Medical University, Dalian 116044, China
2. College of the Humanities and Social Sciences, Dalian Medical University, Dalian 116044, China
3. The First Affiliated Hospital, Dalian Medical University, Dalian 116044, China

Abstract In recent years, the militarization process of artificial intelligence is developing rapidly with the complex and changeable international situation. Starting from the concept of artificial intelligence and its ethical risks, this paper lists the representative ethical issues in the development of military artificial intelligence both in domestic and abroad, including identification technology and moral judgment, responsibility attribution, and disaster caused by the artificial intelligence out of control. This paper analyzes the primary causes of the ethical risks arising from the militarization of artificial intelligence, and puts forward suggestions from three aspects: ensuring the bottom line of avoiding major human disasters caused by military artificial intelligence, continuously improving the accountability gap and technical vulnerability existing in military artificial intelligence, and striving to promote the active discussion of military artificial intelligence topics in the international community, so as to provide ethical review and reference for the application of artificial intelligence to the military.

Keywords artificial intelligence; military application; ethical dilemma; ethical review

（责任编辑　傅雪）