人工智能时代的国家安全：风险与治理

封帅，鲁传颖
（上海国际问题研究院，上海 200233）

[摘要] 人工智能正在成为影响未来社会最重要的技术领域，世界各国纷纷出台指导战略，助力人工智能的创新与发展。然而，人工智能作为一项具有颠覆性的技术，其发展过程本身也蕴含着很大的风险与威胁。人工智能技术发展不仅会导致法律、伦理等方面的问题，也会引发国家安全领域的风险与威胁。本文根据人工智能技术和应用的发展趋势，分析国家在不同领域所面对的安全风险，不仅有助于做好安全防范准备，也有助于处理好安全与发展之间的关系。加强对风险的有效应对，可以更好地保障人工智能技术的发展。在此基础上，国家可以从提升风险意识、完善治理体系、加强监管能力和探求国际合作等多方面来构建风险分析和应对框架，提升国家安全的韧性。

[关键词] 人工智能；国家安全；治理
[中图分类号] TP18;D77　　[文献标识码] A　　[文章编号] 1009-8054(2018)10-0030-20

1 人工智能时代的来临

现代意义上的人工智能研究最早可以追溯到 20 世纪 40 年代。1950 年，英国著名科学家图灵在《计算机器与智能》一文中首次提出了“机器能够思考吗？”这一具有划时代意义的理论问题，并同时提出了测试机器是否拥有智能的方法。

1956 年，在达特茅斯夏季学术研讨会上，研究者们接受了计算机专家约翰·麦卡锡提出的人工智能 (AI) 概念，并将其作为这门新兴学科的正式标签。然而，在随后几十年的时间里，人工智能的发展却并非一帆风顺，虽然在不同时期出现了“专家系统”、“深蓝”等多项具有标志性意义的成果，但由于客观条件的限制，人工智能技术始终无法有效解决人们的现实需求，技术发展无法在产业层面落地，对于现实社会的影响非常有限。

真正的突破出现在 2009–2010 年前后，硬件设备的进步使新一代计算机在运算速度和信息处理能力方面得到大幅提升。互联网产业的发展改变了人类的生活方式，使得网络成为人们获取日常生活数据最集中、最便捷的渠道。移动互联时代的到来，则使来自网络搜索、电子商务、社会媒体、科学研究等不同领域的海量数据迅

速累积，为人工智能的飞跃提供了充足的养分。在已具备强大计算能力与大数据环境的情况下，早在60年代就已经被提出的多层神经网络工具焕发出巨大的生命力，重新成为技术发展的主流路径。

本轮人工智能发展热潮以“深度学习”为核心，该算法以建立套嵌式的多层次模式识别系统组成的“神经”架构为基础，通过组合低层特征形成更加抽象的高层属性、类别或特征，借以发现数据的分布特点。深度学习的出现带来了人工智能算法的跨越，改变了传统符号主义学派以计算机模拟人类认知系统推进人工智能的艰难尝试，让人工智能拥有了从海量且复杂的信息源中提取、识别和构建体系的能力，在那些任务目标明确且相关数据丰富的领域，深度学习算法能够让机器学习新的技能，制定有效策略，从而在短时间内提出超过人类学习能力的问题解决方案。

随着深度学习神经网络成为主流发展方向，人工智能技术很快在现实场景中得以应用。在很短的时间内，人工智能技术就在图像识别、语音识别、机器翻译、自动驾驶甚至棋类竞赛等复杂的应用场景中获得了飞速的进步，基本达到了满足人类日常需求的标准，具备了商业投资的价值，并很快成为资本市场追捧的新热点。

迈过产业化的门槛意味着人工智能技术真正走出了实验室，能够对社会生产和人类生活产生直接影响。更重要的是，本轮人工智能技术的进步证明，机器学习算法能够在具有很多限定性条件的领域比人类做得更好，并且能够通过自我学习不断进步。这一结论事实上打开了人类对于人工智能发展的未来想象空间，人类社会已经能够清晰地听到未来社会匆匆而至的脚步声。

2 人工智能技术与四大领域的安全风险

诚如马克思所言，“科学是历史的有力杠杆，是最高意义上的革命力量”。作为一项具有时代意义的科学思想与技术，人工智能系统能够通过大数据分析和学习理解人类的内在需求，作为创造性的伙伴直接参与到人类改造世界的活动中。它表现出与人类理性思维方式相匹敌的思考能力，在一定程度上改变了人类与技术工具的关系。然而，一切革命性的技术变革都意味着不确定性和风险，人工智能革命也将对经济、政治、军事、社会等领域产生重大冲击。如何在潜在的动荡风险尚未发生时做到未雨绸缪，使技术革命不至于反噬人类本身，是社会科学研究者所肩负的重要责任。在此，本报告分析总结了人工智能技术进步对于社会生活中四大领域所带来的潜在风险，旨在较为全面地展示我们所需面对的挑战。

2.1 人工智能技术对于经济安全的影响

经济安全是国家安全的重要组成部分。特别是在冷战结束之后，随着国际体系的内在逻辑变迁，经济安全问题日益被各国政府和研究者所重视，将其视为国家安全的核心组成部分，并逐渐扩展为一个完整的理论体系。其中金融安全、产业安全、经济信息安全都被视为经济

安全问题的重要组成部分。人工智能技术在发展过程中首先被视为一项具有明确经济目标的技术创新，其成果对于经济体系的影响也最为直接。因此，在尝试讨论人工智能技术所带来的安全风险时，最为基础的便是其对国家层面经济安全的影响。

2.1.1 结构性失业风险

从历史上看，任何围绕着自动化生产的科技创新都会造成劳动力需求的明显下降，人工智能技术的进步也同样意味着普遍的失业风险。

据美国国家科学技术委员会预测，在未来10~20年的时间内，9%~47%的现有工作岗位将受到威胁，平均每3个月就会有约6%的就业岗位消失。[1]与传统基于生产规模下行所导致的周期性失业不同，由新的技术进步所导致的失业现象从本质上说是一种结构性失业，资本以全新的方式和手段替代了对于劳动力的需要。结构性失业的人们在短期内很难重新获得工作，因为他们之前所能够适应的岗位已经彻底消失，而适应新的岗位则需要较长的时间周期。

可以预见的是，主要依赖重复性劳动的劳动密集型产业和依赖于信息不对称而存在的部分服务行业的工作岗位将首先被人工智能所取代。随着人工智能技术在各个垂直领域不断推进，受到威胁的工作岗位将越来越多，实际的失业规模将越来越大，失业的持续时间也将越来越长。①这种趋势的演进，对于社会稳定的影响将是巨大的。

2.1.2 贫富分化与不平等

人工智能技术的进步所带来的另一大经济影响是进一步加剧了贫富分化与不平等现象。

一方面，作为资本挤压劳动力的重要进程，人工智能所带来的劳动生产率提升很难转化为工资收入的普遍增长。在就业人口被压缩的情况下，只有少数劳动人口能够参与分享自动化生产所创造的经济收益。新创造的社会财富将会以不成比例的方式向资本一方倾斜。②

另一方面，人工智能技术对于不同行业的参与和推进是不平衡的。部分拥有较好数据积累，且生产过程适宜人工智能技术介入的行业可能在相对较短的时间内获得较大发展。在这种情况下，少数行业会吸纳巨额资本注入与大量的人才集聚，迅速改变国内产业结构。行业发展不平衡的鸿沟与部分行业大量超额收益的存在将对国家经济发展产生复杂影响。[2]

2.1.3 循环增强的垄断优势

作为一项有效的创新加速器，不断发展成熟的人工智能技术可以为技术领先国家带来经济竞争中的战略优势。人工智能技术的进步需要大量的前期投入，特别是在数据搜集和计算机技术方面的技术积累对于人工智能产业的发

① 相关研究还可参见：Martin Ford. Rise of the Robots: Technology and the Threat of a Jobless Future[M]. New York: Basic Books, 2015; 该书已经出版了中文版，对失业问题的相关论述可见：[英]马丁·福特.机器人时代：技术、工作与经济的未来[M].王吉美、牛筱萌译.北京：中信集团出版社，2015:239-244.

② 根据《乌镇指数：全球人工智能发展报告（2016）》的预测，短期内人工智能的主要应用领域将集中于自动驾驶、医疗健康、安防、电商零售、金融、教育和个人助理7个方面。报告全文见：http://www.199it.com/archives/526338.html.

展至关重要。但各国在该领域的投入差距很大，不同国家在人工智能技术方面的发展严重不平衡。而人工智能技术自身潜在的创造力特性又能使率先使用该技术的国家有更大的机率出现新一轮技术创新。如果这种逻辑确实成立，那么少数大国就会利用人工智能技术实现有效的技术垄断，不仅能够使自己获得大量超额收益，使本已十分严重的全球财富分配两极分化的情况进一步加剧，而且将会随着时间的推移使差距进一步拉大。在这种状况下，处于弱势地位的大部分发展中国家应如何在不利的经济结构中维持自身的经济安全就将成为极具挑战性的课题。

2.1.4 小结

人工智能技术的发展已经深刻地改变着维系国民经济运行和社会生产经营活动的各项基本生产要素的意义。在人工智能技术的影响下，资本与技术在经济活动中的地位获得全面提升，而劳动力要素的价值则受到严重削弱。在传统工业化时代重要的人口红利很可能在新时代成为新型经济模式下的“不良资产”。新的经济体系的重构过程将会引导全球资本和人才进一步流向技术领导国，留给发展中国家走上现代化道路的机遇期将变得更加有限。人工智能技术带来的全球经济结构调整，将促使经济安全问题成为所有发展中国家所面对的共同挑战。

2.2 人工智能技术对于政治安全的影响

人工智能技术对于经济领域的深度影响会自然传导到政治领域，而人工智能技术的特性也容易对现有的政治安全环境产生影响。从议题层面来看，人工智能技术及其背后的大数据和算法能够潜移默化地影响人类行为，直接对国内政治行为产生干扰。从结构层面来看，人工智能所带来的社会经济结构调整，会使资本的权力在政治体系中呈现扩张态势，最终在政治权力分配中获得相应的反映。除此之外，人工智能技术的介入，还将影响国际竞争的内容与形态。因此，对于身处人工智能时代的国家主体而言，如何在变革的条件下有效维护本国的政治安全与秩序稳定，并且提高参与国际竞争的能力，将是所有国家都不得不面对的重要课题。

2.2.1 数据与算法的垄断对于政治议程的影响

技术对于各国国内的政治议程所产生的影响轨迹已经变得越来越清晰，在过去两年中，围绕着 2016 年美国大选而开展的种种政治运作已经越来越明显地展现出拥有数据和技术能够在怎样的程度上影响政治的结果。

剑桥分析公司事件的出现非常清晰地显示出，只要拥有足够丰富的数据和准确的算法，技术企业就能够为竞争性选举制造针对性影响。在人工智能技术的协助下，各种数据资源的积累，使每个接受互联网服务的用户都会被系统自动画像与分析，从而获得定制化的服务。然而，渐趋透明的个人信息本身也就意味着这些信息可以轻易服务于政治活动。正如英国第四频道针对剑桥分析事件所做的评论，“…一只看不见的手搜集了你的个人信息，挖掘出你的希望和恐惧，以此谋取最大的政治利益。”于是，

伴随着技术的不断成熟，当某种特定政治结果发生时，你将难以确定这是民众正常的利益表达，还是被有目的地引导的结果。

在人工智能时代，数据和算法就是权力，这也意味着新的政治风险。这种技术干涉国内政治的风险对于所有国家都普遍存在，但对于那些技术水平相对落后的广大发展中国家来说，这种挑战显然更加严酷。由于缺乏相应技术积累，发展中国家并没有充分有效的方式保护自己的数据安全，也没有足够的能力应对算法所带来的干涉。人工智能技术的进步将进一步凸显其在政治安全领域的脆弱性特征，传统的国家政治安全将面临严峻的考验。

2.2.2 技术进步与资本权力的持续扩张

国家权力的分配方式从根本上说是由社会经济生产方式的特点所决定的，不同时代的生产力水平决定了特定时段最为合理的政治组织模式。威斯特伐利亚体系中的民族国家体制出现，从根本上说正是目前人类所创造的最适宜工业化大生产经济模式的权力分配方式。因此，当人工智能技术所推动的社会经济结构变革逐步深入时，新的社会权力分配结构也会伴随着技术变革而兴起，推动国家治理结构与权力分配模式做出相应的调整。

从当前的各种迹象来看，资本权力依托技术和数据垄断的地位持续扩张将成为新时代国家治理结构调整的重要特征。人工智能技术的研究工作门槛很高，依赖于巨额且长期的资本投入。当前，人工智能研究中最具实际应用价值的科研成果多出自大型企业所支持的研究平台。超级互联网商业巨头实际上掌握了目前人工智能领域的大部分话语权。人工智能领域研究已经深深地打上了资本的烙印，大型企业对于数据资源以及人工智能技术的控制能力正在形成他们实际上的垄断地位。这种力量将渗入当前深嵌于网络的社会生活的方方面面，利用算法的黑箱为大众提供他们希望看到的内容，潜移默化地改变公共产品的提供方式。在人工智能时代，资本和技术力量的垄断地位有可能结合在一起，在一定程度上逐渐分享传统上由民族国家所掌控的金融、信息等重要的权力。资本的权力随着新技术在各个领域的推进而不断扩张，这将成为人工智能技术在进步过程中所带来的权力分配调整的重要特征。

对于民族国家来说，资本权力的扩张本身并非不可接受，大型企业通过长期投资和技术研发，能够更加经济、更加有效地在很多领域承担提供相应公共产品的职能。然而，民族国家能否为资本权力的扩张设定合理的边界则是关系到传统治理模式能否继续存在的重要问题，这种不确定性将成为未来民族国家所面对的普遍性政治安全风险。

2.2.3 技术进步对主权国家参与国际竞争的挑战

人工智能技术进步所带来的另一项重要政治安全风险是使得技术落后的国家在国际战略博弈中长期处于不利地位。

战略博弈是国际竞争活动中最为普遍的形式，参与者通过判断博弈对手的能力、意图、利益和决心，结合特定的外部环境分析，制定出最

为有利的博弈策略并加以实施。[①]由于国际关系领域的战略博弈涉及范围广，内容复杂，各项要素相互累加形成的系统效应（System Effects）实际上已经远远超出了人类思维所能够分析和掌控的范畴，传统意义上国家参与战略博弈的过程更多依赖政治家的直觉与判断。这种类似于“不完全信息博弈”的形态给人工智能技术的介入提供了条件。[②]只要技术进步的大趋势不发生改变，人工智能所提供的战略决策辅助系统就将对博弈过程产生重大影响。[③]

首先，人工智能系统能够提供更加精确的风险评估和预警，使战略决策从一种事实上的主观判断转变为精确化的拣选过程，提升战略决策的科学性。[3-4] 其次，深度学习算法能够以更快的速度提供更多不同于人类常规思维方式的战略选项，并且随着博弈过程的持续，进一步根据对方策略的基本倾向对本方策略加以完善，提升实现战略决策的有效性。[④]最后，在战略博弈进程中，人工智能系统能够最大限度排除人为因素的干扰，提高战略决策的可靠性。[⑤]

以人工智能技术为基础的决策辅助系统在国际战略博弈的进程中将发挥重要作用，技术的完善将使得国际行为体之间战略博弈能力的差距进一步扩大。缺少人工智能技术辅助的行为体将在风险判断、策略选择、决策确定、执行效率，以及决策可靠性等多个方面处于绝对劣势，整个战略博弈过程将会完全失衡。一旦这种情况出现，主权国家将不得不参与到技术竞争中来。而在资本和技术都处于落后一方的中小国家将在国际竞争中处于不利地位，也将面对严重的政治安全风险。

2.2.4 小结

人工智能技术的快速发展所带来的不确定性将直接影响国家的政治安全。它不仅能够直接作用于国内政治议程，通过技术手段对内部政治生态产生短时段的直接干扰，而且会通过国内社会经济结构的调整，在长时段内影响原有政治体系的稳定。在人工智能时代，国内治理格局需要根据经济基础的变化进行调整，作为大工业时代产物的科层制管理体系应该如何适应新时代的要求，将成为影响民族国家国内政治稳定的重要因素。另一方面，人工智能技术的介入和参与还会进一步拉大国家间的战略设计与战略执行能力的差距，技术的潜力一旦得到完全释放，将使得国际竞争格局进一步失衡，处于弱势一方的发展中国家维护自身利益的空间进一步缩减。国际关系行为体之间将呈现出在技术和制度上的系统性差距，发展中国家将面临更加严酷的国际竞争环境。

① 关于战略博弈的分析过程参见：唐世平 . 一个新的国际关系归因理论：不确定性的维度及其认知挑战 [J]. 国际安全研究 ,2014(02):3-41.
② 关于不完全信息博弈问题的基本模式参见：John C. Harsanyi. Games with Incomplete Information Played by “Bayesian” Players, Part I: The Basic Model[J].Management Science, 1967, 14(03):159-182.
③ 参见：Kareem Ayoub, Kenneth Payne.Strategy in the Age of Artificial Intelligence[M].The Journal of Strategic Studies, 2016,39(5-6):793-819.
④ Kareem Ayoub, Kenneth Payne. Strategy in the Age of Artificial Intelligence[M].The Journal of Strategic Studies, 2016, 39(5-60:808.
⑤ 参见：Stephen P. Rosen.War and Human Nature[M]. Princeton, NJ: Princeton University Press, 2005:27-70.

2.3 人工智能技术对于军事安全的影响

人工智能技术本身并不是军事武器，但它天然与军事安全领域的所有问题都存在千丝万缕的联系。从人工智能技术诞生之日起，如何将其有效应用于军事领域就已被纳入所有技术先进国家的考虑范围之内。[①]这是因为国家的军事行为与公司等经济组织的商业行为拥有相似的逻辑，都要求建立一个有效的系统，以便在竞争性过程中获得胜利。整个过程中同样包含快速获取信息、快速处理信息、做出决策与执行决策等过程。而随着人工智能技术的成熟，它将会被越来越广泛地应用于军事领域，武器系统、军事策略、军事组织，甚至战争的意义可能会发生深刻改变，人类社会也有可能在进入人工智能时代之后迎来一个不同的军事安全环境。[②]

2.3.1 完全自主性武器的广泛应用将带来巨大的军事伦理问题

人工智能技术不是武器，但能够成为武器性能提升的助推器。一方面，人工智能技术的介入，使大量无人作战武器参与作战成为可能。当前，无人机、无人地面车辆、无人潜航设备已经广泛应用于军事领域，而各国军事部门对于机器人作战系统的兴趣也是有增无减。利用深度学习算法，智能化武器可以在虚拟环境中得到武器操控的基本能力，随后在现实环境中广泛获取数据，并根据数据反馈不断提升战斗能力，学习执行各种战斗命令，最终实现有效应用于复杂的战场态势。

另一方面，随着人工智能技术的发展，算法的更新可以赋予智能武器新的角色与行动逻辑。以智能化无人机为例，利用人工智能技术，无人机已不仅是执行定点清除等特殊任务的执行者，更成为情报搜集、目标定位、策略制定和行动发起的协调平台，担负起前沿信息节点和策略制定等重要任务。此外，人工智能技术的成果同样可以应用于对于各种智能化武器的训练过程中。智能化武器的规模越大，其在战斗中相互协调的优势就越容易发挥出来。通过共同的算法进行“训练”的大批量智能化武器可以协调行动，有助于其最大限度地优化其作战策略，并且根据战场形势和作战目标进行灵活调整，最大限度地获得战场优势。[③]

然而，武器系统的快速进步也为国家的军事行为带来了严重的伦理问题。随着技术的进步，完全自主的致命性武器系统能够做到主动识别和选择目标，确定拟对目标施加的武力级别，并在特定的时间和空间范围内对目标实施规定的武力。但自主武器是否有权力在没有人为干涉的情况下自主决定对于目标的杀伤，仍

① 在人工智能技术诞生之初，美国军方就自动化和智能化武器产生的浓厚的兴趣，相关信息可参见：Allan M. Din (ed.). Arms and Artificial Intelligence: Weapon and Arms Control Applications of Advanced Computing[M].New York: Oxford University Press, 1988; Jeffrey L. Caton. Autonomous Weapons Systems: A Brief Survey of Developmental, Operational, legal and Ethical Issues[EB/OL].Strategic Studies Institute, U.S. Army War College.(2015-11). http://www.strategicstudiesinstitute.army.mil/pdffiles/PUB1309.pdf.

② Vincent Boulanin, MaaikeVerbruggen.Mapping the Development of Autonomy in Weapon Systems[M].Sweden Stockholm International Peace Research Institute, 2017:16-17.

③ Vincent Boulanin, MaaikeVerbruggen.Mapping the Development of Autonomy in Weapon Systems[M]. Sweden: Stockholm International Peace Research Institute, 2017:27-29; Peter Singer, Wired for War: The Robotics Revolution and Conflict in the Twenty-first Century[M]. London: Penguin, 2009:124-125.

然是人类伦理领域的一个尚无答案的问题。人类社会的运行要建立在很多具有共识性的伦理基础之上，即使是军事行为也有很多明确的国际法规范。然而，这些法律规范都立足于人类之间的战争行为，对于智能化武器的规范尚未形成。特别是处于弱势一方的军事组织，在无法通过消灭有生力量的方式制止对方战争行为的情况下，是否有权利对于对方城市平民发动袭击，迫使对方停止侵略行为？如果这种行为能够被接受，那么军事行动的合法性界限到底在哪里？在这些问题得到有效解决之前，一旦在现实战场上出现智能化武器自主决定对于人类的大规模杀伤，人类社会的伦理原则就将面临重大考验。

2.3.2 更加有效的作战体系的出现很可能触发新一轮的军备竞赛

人工智能技术的进步和智能化武器的发展，可以使人工智能时代的作战体系逐渐趋向去中心化的动态网络结构。由于智能化武器本身有承载作战关键节点的功能，且相互之间能够实现数据和策略共享，因此，在战争过程中能够做到相互取代，从而避免了因为关键节点被攻击而导致整个作战系统失效的结果。同时，人工智能具有更加全面高效搜集战场信息的能力，能够利用智能系统重新构筑战场形态，实现对战场真实情况最大限度的模拟。在人工智能技术的推动下，在军事安全领域能够出现更加有效的作战体系。

事实上，人工智能拥有两个人类无法比拟的优势，其一，人工智能系统可以快速处理战场信息，具有人类所不具备的快速反应能力。其二，人工智能系统具有多线程处理能力，可以同时处理军事行动中同时发生的多项行动，并且提出人类思维模式所无法理解的复杂策略。[①]速度是现代战争中的重要优势，在现代战争信息超载的情况下，成熟的人工智能系统的反应速度和策略安排都将远远超过人类体能的极限。技术的影响将加剧常规军事力量对抗的不平衡状态，缺少人工智能技术辅助的武装力量将越来越难以通过战术与策略弥补战场上的劣势。常规对抗将不再是合理的战略选项，不对称战争将成为这两种力量对抗的主要方式。

人类既有的历史经验多次验证了任何科技革命的出现都会使率先掌握新科技的国家与其他国家之间的力量差距进一步扩大。作为人类科技史上最新的力量放大器，人工智能在军事领域已经展现出明显超越人类的能力与持续发展的潜力。一旦技术发展成熟，这种差距已经很难用数量堆砌或策略战术加以弥补，应用人工智能的国际行为体在军事行动中很难被尚未使用人工智能技术的对手击败，国际主体间的力量鸿沟变得更加难以跨越。面对这样的技术变革浪潮，所有具有相应技术基础的大国必然会千方百计地获取相关技术，一场以人工智能技术为核心的新的军备竞赛恐怕很难避免。

2.3.3 人工智能技术会大大降低战争的门槛

在现代国际体系中，战争被普遍视为国际政治行为中的极端手段。巨大的经济成本

① Li Yingjieet, al.RTS Game Strategy Evaluation Using Extreme Learning Machine[J].Soft Computing, 2012, 16(09):1623-1637.

与伤亡所造成的国内政治压力实际上给战争设置了较高的门槛。然而，随着人工智能技术的介入，战争行动的成本与风险都有明显下降的趋势。

一方面，人工智能技术的介入将能够有效节约军事行动的成本。智能化武器的使用可以有效节约训练过程的时间和人力成本。无人作战武器的训练多依赖于相对成熟的深度学习算法，在初始训练结束后，可以快速复制到所有同类型无人作战武器上，完成作战武器的快速培训过程。最大限度地节省了人类武器操控者需要对所有个体重复培训的人力和物力成本，而且可以从整体上做到所有武器操控的同步进步。从长时段效果来看，这更是一种更加经济、更加有效的作战训练方式。由于算法与数据的可复制性，部分武器的战损对于整体作战效能的影响将大大降低。即使在实际战斗中出现战损情况，其实际损失也要明显小于传统作战武器。

另一方面，传统战争模式中最为残酷的一面是战争导致的人员伤亡，这也是现代社会战争行为最为严重的政治风险。而智能化武器的广泛应用实际上减少人类直接参与战斗的过程，人与武器实现实质性分离，将战争活动在很大程度上转变为利用无人武器系统的任务。从而有力地规避了大量伤亡所导致的政治风险。在传统的战争形态中，由于人类的深度参与，战争的双方都有较大的可能出现重大伤亡，这是战争的不确定性所决定的。在现代政治体系中，战争所导致的大量本国人员伤亡会在国内政治领域形成重要的社会压力，客观上增加了大国发动战争的顾虑，提升了战争的门槛。然而，随着智能化武器的广泛使用，人员伤亡能够大大减少，政治风险极大降低。这种情况实质上鼓励大国减少自我约束，更多采取进攻性的行动来达到相应的目的，也会对国际安全形成新的不稳定因素，客观上为大国之间的技术军备竞赛提供了额外的动力。

2.3.4 人工智能技术给网络安全问题带来的重大风险

网络安全本身就是具有颠覆性、杀手锏性质的领域，人工智能的应用将会进一步放大网络安全在进攻和防御方面的作用，从而使得强者愈强。同时，人工智能在网络攻击行动和网络武器开发中的应用也会带来很大的安全风险。这种风险主要表现在对自主选择目标的攻击是否会引起附带的伤害，是否会超出预设的目标从而导致冲突升级。在现有网络领域的冲突中，各方在选择目标和采取的破坏程度时都会非常谨慎，避免产生不必要的伤害以及防止冲突发生。但是人工智能网络武器的使用是否能够遵循这一谨慎，能否将更多在网络安全之外的因素纳入到攻击目标的选择和攻击程度的决策上，仍然存在疑问。因此，自主攻击的网络武器开发应当被严格限制在特定的环境之下，并且精确地开展测试。

另一方面，自主攻击网络武器的扩散将会对网络安全造成更加难以控制的危害。近年发生的网络武器泄露已经给国际安全造成了严重威胁，类似于 WannaCry 和 NotPetya 这样源自于美国国家安全局武器库中网络武器泄露再次开发而成的勒索病毒给国际社会带来了几百亿美

元的经济损失和重大的公共安全危害。如果更具危害的自主性网络武器一旦泄露，其给网络安全带来的威胁将会更加严重。试想如果恐怖主义集团获得了可以自动对全球各个关键基础设施发动攻击的网络武器，那么将会对全球网络安全造成严重危害。因此，自主网络武器需要有严格加密和解密的规定，并且还应当具有在泄露后自我删除、取消激活等功能。

2.3.5 技术本身的安全问题与技术扩散对于全球安全的威胁

人工智能技术的介入能够使军事武器的作战效能提升，同时推动成本逐步下降，两方面优势的同时存在将使得对智能化武器的追求成为各大国的合理选择，但这并不意味着人工智能技术已经完全解决了可靠性的问题。从目前情况看，人工智能技术本身的安全问题与技术扩散风险仍然不可忽视。

一方面，技术本身仍存在潜在的安全问题。算法与数据是人工智能技术发展最重要的两项要素，但这两项要素本身都蕴含着潜在的安全风险。算法是由人编写的，因此，无法保证程序完全安全、可靠、可控、可信。而从数据角度来看，人工智能依赖大数据，同时数据的质量也会影响算法的判断。军事数据的获取、加工、存储和使用等环节都存在着一定的数据质量和安全风险问题。军队的运作建立在可靠性的基础之上，而人工智能技术本身存在的不确定性会为全球军事安全带来考验。

另一方面，人工智能技术的扩散给全球安全带来了威胁。伴随着人工智能武器的开发，国际社会面临将面临严峻的反扩散问题的挑战。恐怖主义组织以及部分不负责任的国家有可能利用各种途径获得人工智能武器，并对国际安全和平产生威胁。人工智能从某种意义上而言，也是一种程序和软件，因此，它面临的扩散风险要远远大于常规武器。经验表明，类似于美国国家安全局的网络武器库被黑客攻击，并且在暗网进行交易，最后被黑客开发为勒索病毒的案例也有可能在人工智能武器领域重现。如何控制人工智能技术扩散所带来的风险将成为未来全球军事安全的重要议题。

2.3.6 小结

人工智能技术在军事领域的深度介入，是核武器发明以来全球军事领域所出现的最重要的技术变革之一。[①]以深度学习为标志的人工智能技术可以增强信息化作战系统的能力，这是改变战争形态的基础。智能化武器的出现在理论上能够为国家提供低成本和低风险的军事系统，能够再次在短时间内放大主体间军事力量的差距，拥有人工智能技术的国家将具有全面超越传统军事力量的能力，使对方原本有效的伤害手段失效。新的不平衡状态可能会造成重大的伦理问题，而中小国家则不得不面对更加严酷的军事安全形势。如果这种状况不能得到

① Greg Allen, Taniel Chan.Artificial Intelligence and National Security, Intelligence Advanced Research Projects [M].Activity(IARPA), Belfer Center, Harvard Kennedy School, 2017:15−18.

有效管控，大国将陷入新一轮军备竞赛，而中小国家则必然会寻找相关军事技术的扩散或新的不对称作战方式，以便维持自己在国际体系中的影响能力。

2.4 人工智能技术对于社会安全的影响

作为新一轮产业革命的先声，人工智能技术所展现出来的颠覆传统社会生产方式的巨大潜力，以及可能随之而来的普遍性失业浪潮将持续推动着物质与制度层面的改变，也持续地冲击着人们的思想观念。面对剧烈的时代变革与动荡，世界各国都面临着法律与秩序深度调整、新的思想理念不断碰撞等问题。变革时期的社会安全问题也将成为各国不得不面对的重要挑战，新的思想与行动最终将汇集形成具有时代特征的社会思潮，对国家治理方式产生重要影响。

2.4.1 人工智能技术带来的法律体系调整

人工智能技术在社会领域的渗透逐渐深入，给当前社会的法律法规和基本的公共管理秩序带来了新的危机。新的社会现象的广泛出现，超出了原有的法律法规在设计时的理念边界，法律和制度产品的供给出现了严重的赤字。能否合理调整社会法律制度，对于维护人工智能时代的社会稳定具有重要意义。针对人工智能技术可能产生的社会影响，各国国内法律体系至少要在以下几个方面进行深入思考：

（1）如何界定人工智能产品的民事主体资格

尽管目前的人工智能产品还具有明显的工具性特征，显然无法成为独立的民事主体，但法律界人士已经开始思考未来更高级的人工智能形式能否具有民事主体的资格。事实上，随着人工智能技术的完善，传统民法理论中的主体与客体的界限正在日益模糊，学术界正在逐步形成“工具”和“虚拟人”两种观点。所谓“工具”，即把人工智能视为人的创造物和权利客体；所谓“虚拟人”是法律给人工智能设定一部分“人”的属性，赋予其能够享有一些权利的法律主体资格。这场争论迄今为止尚未形成明确结论，但其最终的结论将会对人工智能时代的法律体系产生基础性的影响。

（2）如何处理人工智能设备自主行动产生损害的法律责任

当人工智能系统能够与机器工业制品紧密结合之后，往往就具有了根据算法目标形成的自主性行动能力。然而，在其执行任务的过程中，一旦出现对于其他人及其所有物产生损害的情况，应如何认定侵权责任就成了一个非常具有挑战性的问题。表面上看，这种侵权责任的主体应该是人工智能设备的所有者，但由于技术本身的特殊性，使得侵权责任中的因果关系变得非常复杂。由于人工智能的具体行为受算法控制，发生侵权时，到底是由设备所有者还是软件研发者担责，很值得商榷。

（3）如何规范自动驾驶的法律问题

智能驾驶是本轮人工智能技术的重点领域，借助人工智能系统，车辆可以通过导航系统、传感器系统、智能感知算法、车辆控制系统等智能技术实现无人操控的自动驾驶，从而在根本上改变人与车之间的关系。

无人驾驶的出现意味着交通领域的一个重

要的结构变化，即驾驶人的消失。智能系统取代了驾驶人成为交通法规的规制对象。那么一旦出现无人驾驶汽车对他人权益造成损害时，应如何认定责任，机动车所有者、汽车制造商与自动驾驶技术的开发者应如何进行责任分配。只有这些问题得以解决，才能搭建起自动驾驶行为的新型规范。

归结起来，人工智能技术对于社会活动所带来的改变正在冲击着传统的法律体系。面对这些新问题和新挑战，研究者必须未雨绸缪，从基础理论入手，构建新时代的法律规范，从而为司法实践提供基础框架。而所有这些都关系到社会的安全与稳定。

2.4.2 思想理念的竞争性发展态势

随着人工智能技术的发展和进步，特别是“机器替人”风险的逐渐显现，人类社会逐渐针对人工智能技术也将逐渐展示出不同的认知与思想理念。不同思想理念之间的差异与竞争反映了社会对于人工智能技术的基本认知分歧。同时，不同思想理念所引申的不同策略与逻辑也将成为未来影响人类社会发展轨迹的重要方向。

（1）第一种可能广泛出现的思想理念是：保守主义

事实上，在每一次工业革命发生时，人类社会都会出现对于技术的风险不可控问题的担忧，人工智能技术的进步也概莫能外。在深度学习算法释放出人工智能技术的发展潜力之后，在很多领域的人工智能应用系统都仅仅需要很短的学习时间，便能够超越人们多年所积累的知识与技术。人类突然意识到，自己曾经引以为傲的思维能力在纯粹的科学力量面前显得是那样微不足道。更严重的是，深度学习算法的“黑箱”效应，使人类无法理解神经网络的思维逻辑。人类对未来世界无法预知和自身力量有限而产生的无力感所形成的双重担忧，导致对技术的恐惧。这种观念在各种文艺作品都有充分的表达，而保守主义就是这种社会思想的集中反映。

在他们看来，维持人工智能技术的可控性是技术发展不可逾越的界限。[①]针对弱人工智能时代即将出现的失业问题，保守主义者建议利用一场可控的“新卢德运动”[②]延缓失业浪潮，通过政治手段限制人工智能在劳动密集型行业的推进速度，使绝对失业人口始终保持在可控范围内，为新经济形态下新型就业岗位的出现赢得时间。这种思路的出发点在于尽可能长地维护原有体系的稳定，以牺牲技术进步的速度为代价，促使体系以微调的方式重构，使整个体系的动荡强度降低。

① 关于超人工智能主要特征可参见：Nick Bostrom.Superintelligence: Paths, Dangers, Strategies[M].Oxford: Oxford University Press, 2014.

② 卢德运动（Luddism）是19世纪初英国工业革命时期，传统纺织业者捣毁机器的群众运动。20世纪90年代之后，新一代反对现代技术的哲学思潮逐渐出现，因为出于对自动化、数字化负面影响的担心，他们希望限制新技术的使用，由于同早期卢德运动的思想渊源相近，因此被称为新卢德运动（Neo-Luddism）。在弱人工智能时代，预计新卢德运动将获得更多的支持，有可能成为一股重要的社会思潮。关于新卢德运动的更详细介绍可参见：Steven E. Jones.Against Technology: From the Luddites to Neo-Luddism[M].New York: Routledge, 2006.

然而，在科技快速发展的时代，任何国家选择放缓对新技术的研发和使用在国际竞争中都是非常危险的行为。人工智能技术的快速发展可以在很短的时间内使得国家间力量差距被不断放大。信奉保守主义理念的国家将在国际经济和政治竞争中因为技术落后陷入非常不利的局面，这也是保守主义思想的风险。

（2）第二种可能广泛出现的思想理念是：进步主义

这种观点的理论出发点在于相信科技进步会为人类社会带来积极的影响，主张利用技术红利所带来的生产效率提升获得更多的社会财富。进步主义体现了人类对于人工智能技术的向往，这一思想理念高度评价人工智能所引领的本轮工业革命的重要意义。他们解决问题的逻辑是要通过对于制度和社会基本原则的调整，充分释放人工智能技术发展的红利，在新的社会原则基础上构建一个更加适应技术发展特性的人类社会。

在进步主义者看来，人工智能技术所导致的大规模失业是无法避免的历史规律，试图阻止这种状态的出现是徒劳的。维持弱人工智能时代社会稳定的方式不是人为干预不可逆转的失业问题，而是改变工业化时代的分配原则。利用技术进步创造的丰富社会财富，为全体公民提供能够保障其保持体面生活的收入。[①]最终实现在新的分配方式的基础上重新构建社会文化认知，形成新时代的社会生活模式。

进步主义思想的主要矛盾在于，它的理论基础建立在人工智能技术能够快速发展并能够持续创造足够丰富的社会财富的基础上，从而满足全球福利社会的需求。然而，人工智能技术的发展历史从来不是一帆风顺，从弱人工智能时代到强人工智能时代需要经历多久，至今难有定论。一旦科技进步的速度无法满足社会福利的财富需求，进步主义所倡导的新的社会体系的基础就将出现严重的动摇，甚至会出现难以预料的社会剧烈动荡。

2.4.3 小结

变革必然意味着风险，风险则会带来社会安全的挑战。人工智能技术的发展和进步能够直接作用于经济、政治等多个领域，也对于社会结构将产生深远影响。面对技术所带来的社会安全风险，我们既需要积极调整法律规制体系，努力维持社会稳定，又要在思想层面上对本轮社会变迁进行深层次的思考。虽然面对奔涌而来的人工智能浪潮，不同的思想理念展现的应对路径具有明显的分歧，但无论怎样，我们在思考人类与人工智能技术的关系时，应该始终坚信，人工智能是人类的造物，是人类知识与理性的伟大结晶。人工智能可能给世界带来的威胁远远不及那些人类自己可能创造的恶。我们应该以冷静而客观的态度理解和思考人工智能技术对于社会的影响，在处于变革中

① 关于“无条件基本收入”的相关问题，参见：Phillipe Van Parijs. Basic Income: A Radical Proposal for a Free Society and a Sane Economy[M].Boston: Harvard University Press, 2017; Karl Widerquist (ed.). Basic Income: An Anthology of Contemporary Research[M].Hoboken, NJ: Wiley-Blackwell, 2013.

的并且更趋不平等的世界创造更加稳定、更加合理、更加体现人类文明与尊严的体系与制度。

3 人工智能在国家安全领域的风险应对

目前世界各国政府在人工智能领域的主要关注点是推动发展，对其所蕴含风险则准备不足。如上文所述，在人工智能时代，世界各国在国家安全领域面临的各项风险是相当严峻的。从全球层面来看，各国所具备的应对举措还存在较大的缺点。结合人工智能技术特点和应用发展趋势中可能引发的各项潜在风险，我们认为，国家可以从风险意识提升、治理体系建设、加强监管能力和国际合作等多个方面来构建风险分析和应对框架，以提升国家安全的韧性。

3.1 提升风险防范意识

提升风险防范意识是应对人工智能时代国家安全风险的重要起点。相对于其他领域的安全风险，人工智能在国家安全领域的风险具有系统性、不确定性、动态性等特点。此外，人工智能是一个新的风险领域，既有的安全治理经验很少，人们很难从过去的经验中吸取教训。因此，无论是风险的识别、预防和化解都是一项全新的挑战。建立相应的风险感知、识别、预防和化解能力框架是现阶段应对人工智能风险的当务之急。

3.1.1 感知风险意识

国家在发展和应用人工智能技术过程中，应当重视提高对技术以及应用的风险意识。由于人工智能技术的复杂性，企业常常处于技术创新和发展的前沿；而国家在某种程度上远离技术的前沿，对技术的感知存在一定的滞后，并且往往是在事件发生之后才被动地做出反应，这样就会错过最佳的干预时期。为了建立主动应对的能力，国家首先需要提高对于行业和领域的风险意识，避免由于风险意识不足导致的危机。

例如，在总体国家安全观以及其包含的政治安全、国土安全、军事安全、经济安全、文化安全、社会安全、科技安全、信息安全、生态安全、资源安全、核安全等在内的 11 个安全领域中高度重视人工智能发展可能带来的积极和消极影响。特别是在涉及到政治安全、军事安全、经济安全、信息安全、核安全等领域，人工智能所包含的风险已经开始显现，但相应的风险意识并没有跟上风险的威胁程度。[①]在这些重点领域和行业，应当把提升风险意识作为制度性工作。

提升风险意识需要国家密切关注人工智能技术和应用的发展，通过系统思维对其可能在重要安全领域带来的风险进行深度思考。提升意识有助于后续的风险识别、预防和化解的过程，增加国家和社会对风险的重视程度，从而加大资源的投入与人才的培养。

3.1.2 识别风险能力

识别潜在的风险是加强危机应对的重要组

① Gregory C. Allen,Taniel Chan. Artificial Intelligence and National Security[M].Cambridge: Harvard University, 2017.

成部分，但它又是具有挑战性的工作。人工智能距离广泛应用还有一段距离，相应的风险还在逐步地显现，在缺乏足够的案例的情况下，建立对风险的识别能力是一种具有前瞻性和挑战性的工作。人工智能是一项通用性的技术，有很多方向和突破点，这加大了风险识别的难度。总体而言，主要遵循着技术突破——应用创新——新的风险这样一个过程。识别风险的阶段越早，对于风险的控制就越容易。已有的案例和技术发展的趋势表明，人工智能所带来的风险程度高，往往还是具有一定系统性特征，对国家安全所造成的威胁程度较大。对于国家而言，识别人工智能的风险能力建设是一项长期的工作，需要建立跨学科的知识背景以及相应的跨部门协作机制，在政治安全、经济安全、军事安全、社会安全等领域建立相应的风险识别机制，加强相应的能力建设。

3.1.3 风险预防能力

风险预防是指对已经识别的风险和未能识别的风险进行预防。对已经能够识别的风险领域，应当根据自身的脆弱性，制定相应的预案，并且寻求风险降级的方法。对于未能识别的风险，则需要投入更多的精力，制定相应的规划，评估处置措施。在国家安全领域建立风险预防能力对于政府部门的动员能力有很高的要求。在很多风险的预防问题上，政府都缺乏足够的经验，缺乏成熟的应对机制，但是却需要政府部门能够快速地应对，及时的制定相应的风险预防计划。

3.1.4 风险化解能力

风险化解的能力，最终决定国家在人工智能时代应对国家安全风险的结果。风险意识提高、识别能力加强和建立预防能力都是增加风险化解能力的关键。但是，最终如何化解风险还取决于多方面的能力要素构建。人工智能所具有的跨领域特征，要求首先构建相应的跨部门协调机制。[①]人工智能所展现的跨国界特征，则要求建立起相应的国际合作机制。总体而言，如果化解人工智能的风险，就需要持续的关注和不断的加强能力建设，这对国家提出了更高的要求。

最后，风险化解一项系统性工程，它并非是要减少和限制人工智能的发展，相反，它是建立在对人工智能技术、应用、影响等多个维度的精确理解基础之上，在发展与安全之间取得平衡的一种能力。风险管控越有效，人工智能发展的空间越大，国家竞争力就越强。因此，提升国家安全领域的人工智能风险意识以及建立相应的管控机制，是保障其未来发展的关键。

3.2 构建安全治理体系

在提升风险意识和加强风险应对的同时，国家还应当主动加强人工智能安全治理体系的建设，将更多的利益攸关方纳入到治理体系当中，从技术、产业、政策、法律等多个方面建立安全保障体系。

人工智能是知识密集型和资本密集型的领

① Miles Brundage, ShaharAvin, Jack Clark, et alia. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation[M].Cornell: arXiv Archive, Cornell University Library, 2018.

域，人工智能技术的开发者，包括算法和数据领域的专家以及应用开发领域的人才大多分布在企业、研究机构当中。因此，构建多方参与的治理体系是推动人工智能治理的主要方式。①其中，政府作为监管部门，加强与社群和私营部门之间的互动，一方面可以掌握技术和应用发展的趋势和方向，另一方面也可以帮助社群和私营部门更好地理解政府的关切，从而避免因为忽视和误判引起的不必要安全风险。

治理不同于管理，很大程度上治理是一种开放、多方参与的过程。很多情况下，治理是自下而上的，强调技术导向的治理方式。它与传统的自上而下、只有政府参与的管理模式存在很大的不同。目前，以多利益攸关方为主要治理模式的机制正在人工智能领域出现。②例如，在人工智能的安全与管理领域，正在形成新的以社群为中心的治理机制，IEEE《以伦理为基础的设计》《阿西洛马倡议》是其中的代表。而国际电信联盟这样的政府间组织也在通过多利益攸关方模式发起《向善的人工智能倡议》。[5]

政府应与其他的利益攸关方之间加强交流合作，通过相应的互动机制帮助各方更好地理解国家安全领域的风险，以及共同制定相应的风险管控机制；避免由于知识与政策之间的鸿沟影响技术社群、私营部门与政府之间的沟通障碍，导致风险出现以及相应的风险管控措施无法出台。

3.3 加强监管能力建设

从国家安全角度来看，政府是应对风险最主要的责任人，加强政府的监管能力是降低人工智能风险、保障技术与产业长足发展的关键。从政府角度而言，建立人工智能的技术标准体系、应用开发的规范体系以及建立与人工智能时代相适应的法律体系是保障监管能力的关键。

3.3.1 技术标准体系

人工智能的技术标准对于技术本身是否存在漏洞、符合相应的安全要求具有重要的作用。技术上的漏洞有可能导致人工智能的系统被黑客或其他不法分子利用，从而危害国家安全。例如很多智能音箱所暴露出来对个人隐私的侵犯，以及无人驾驶汽车由于图像识别能力不足导致的车祸等，都暴露出了人工智能还存着很多的不完善的地方，有可能会引发新的安全风险。对于这些缺陷，应当不断提高技术标准予以克服。

技术标准由技术社群主导建立，直接面向工程师和开发人员，对于提高人工智能应用的安全性具有重要作用。例如，作为主要的人工智能技术社群，IEEE 设立了 IEEEP7000 标准工作组，设立了伦理、透明、算法、隐私等 10 大标准工作组，通过国际化的标准了影响整个技术社群。[5] 2018 年 1 月，中国国家标准化管理委员会宣布成立国家人工智能标准化总体组、专家咨询组，负责全面统筹规划和协调管理我国人工智能标准化工作。[6] 标准虽然是面向开发

① Joseph S. Nye.The Regime Complex for Managing Global Cyber Activities[D].Harvard University, 2014.
② ITU.AI for Good Global Summit Report[EB/OL].Geneva.(2017-06-30).https://www.itu.int/en/ITU-T/AI/Pages/201706-default.aspx.

者，但是它作为技术人员所应遵循的基本规范，对于国家安全而言具有重要作用。

3.3.2 程序规范体系

在标准之上，国家还应当在不同的部门制定相应的规范流程，确保对安全风险的控制。以军事领域为例，无论是半自主和全自主的武器系统，其设计应允许指挥官和作战人员在使用武力方面作出适当的人为判断。在决定是否使用人工智能武器时，应当制定明确的规范和流程，避免造成不可预测的后果。①同时，还应当加强对人工智能武器的安全性，避免武器泄露或者随意转让相关技术。无论是无意或者有意的扩散都会增加军备竞赛的风险，威胁国际安全体系。

同时对于人工智能军事领域的使用，也应当有一套严格的流程，确保“合法使用”，即负责授权使用、指挥使用或操作自动化和半自主武器系统的人必须遵守战争法、其他国际条约、武器系统的安全规则以及适用的交战规则（ROE）。

3.3.3 法律法规体系

为应对人工智能时代的安全问题，国家应当建立相应的法律、法规体系。国际和国内法律界对此展开了激烈讨论。有学者认为，“智能革命的出现，对当下的伦理标准、法律规则、社会秩序及公共管理体制带来一场前所未有的危机和挑战。它不仅与已有法律秩序形成冲突，凸显现存法律制度产品供给的缺陷，甚至会颠覆我们业已构成的法律认知。”[7]具体而言，人工智能已经在法律资格的主体、致人伤害的责任认定等多个方面提出了现实的问题。

从国家安全角度来看，人工智能的法律、法规所包含的内容更加丰富。从军事领域看，在应对人工智能武器攻击时，如何从国际法角度去认定攻击的性质和攻击的溯源，并应采取何种形式的反击和应对举措？在经济领域，如何去规范人工智能造成的金融系统安全问题？如何明确相应的责任，对相关的企业和人员进行处罚？在政治安全领域，对于类似“剑桥分析”的大数据公司和互联网社交媒体平台，应当如何制定相应的法律法规，禁止其通过人工智能对政治安全进行干扰和破坏？

3.4 提升国际合作水平

人工智能所带来的问题具有全球属性，它不是某一国家或者某一领域的治理问题。从技术本身来看，人工智能算法与安全问题是人类共同面临的挑战，具有普适性；从应用角度来看，各国在人工智能的发展和应用上对国家安全造成的威胁是跨国界的；从体系角度来看，人工智能对于地缘经济、地缘安全的颠覆性影响，冲击甚至重塑着现有的国际政治体系，从而影响体系中每一个行为体的国家安全。因此，加强人工智能领域的国际合作对于应对国家安全风险具有重要作用。

3.4.1 国际法

联合国高度关注人工智能对于国家安全的

① US DOD, Autonomy in Weapon Systems, No. 3000.09, Nov 21, 2012.

影响，经过2014年和2015年的两次非正式会议后，在2016年12月16日关于特定常规武器公约的联合国会议上成立了致命性自主武器系统（LAWS）政府专家组（GGE）。该小组的任务是研究致命性自主武器领域的新兴技术，评估其对国际和平与安全的影响，并为制定相应的国际法和国际治理提供建议。[①]2018年9月在日内瓦召开的联合国常规武器公约的讨论中，各方就制定禁止人工智能驱动的致命性完全自主武器的条约开展讨论。全球约有26个国家支持全面禁止人工智能武器，而绝大多数国家还在观望，所以这次会议并未达成共识。[②]但是，在规则缺失和大国战略竞争的背景下，军事领域人工智能的发展带来的风险和威胁在不断增加，出台国际法对于减小人工智能安全风险至关重要。

技术强国与弱国之间存在不同的观点是导致相应的国际法难以出台的重要原因。技术弱国认为应当完全禁止致命性自主武器的开发使用，技术强国则持相反意见，认为开发致命性自主武器可以降低人员损伤，有利于打击恐怖主义和维护国家安全，并且很多系统已经在战场上进入实战。军事是推动技术进步的重要因素，互联网就是由美国军方所发明，同样，人工智能的技术发展背后也有军事因素在强力推动。但是，人工智能的军备竞赛也非技术之福，特别是致命性自主武器的扩散会造成更为严重的后果。因此，从联合国层面制定相应的国际法，并且促成大国之间在发展致命性自主武器上达成一定的军控条约是当务之急。

当前，在国际法不明确的情况下，各国应克制在军事领域使用人工智能武器。联合国政府专家组也应考虑对国际法此前未曾预见的情况追加法律限制，并以道德或伦理为由尽量减少对人和平民的伤害。更进一步的目标包括管理以下风险，如使用武力的门槛降低、意外导致的伤害、预期之外的矛盾升级以及军备竞赛和扩散。

3.4.2 双边及多边合作

从联合国角度来达成一项谈判可能需要十年甚至更长的时间，鉴于越来越多的风险，大国之间应当及早就人工智能在军事领域应用可能带来的潜在风险开展对话，及早启动相应的军控进程，短期内促成技术强国在谋求战略优势和国际安全体系稳定之间的妥协。同时，各国政府在面对人工智能时代的国家安全问题上面临着共同的威胁和挑战，加强彼此之间的合作是应对威胁的解决之道。

在军事领域，大国之间可以就人工智能的发展以及可能触发的军备竞赛问题开展对话，避免由于过度、过快的开发军事应用而引起的伦理问题，以及对国际安全体系的稳定造成冲击。可以预见，人工智能在军事领域的应用会

① UNICRI.CBRN National Action Plans: Rising to the Challenges of International Security and the Emergence of Artificial Intelligence[EB/OL].(2015-10-07).http://www.unicri.it/news/article/CBRN_Artificial_Intelligence.

② UNIDIR.The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches[EB/OL].Geneva: 2017. http://www.unidir.org/files/publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-concerns-characteristics-and-definitional-approaches-en-689.pdf.

加速发展，存在的问题也会越来越多。特别是各国之间对于如何看待人工智能武器，如何进行有效互动还存在很大的认知差距。这种差距的存在，导致各国在冲突发生时缺乏有效的应对手段。因此，各国之间应当就人工智能的军事安全问题开展对话，加强管控，增强政策透明度、建立信任措施，降低相应的军事冲突的风险，并且在冲突发生时能够采取有效的危机管控和冲突降级措施。

在政治安全、经济安全、社会安全领域，许多国家也都在积极开展各种实践活动，相应的做法之间有很大的相互借鉴之处。各国可以就政策实践、信息共享、最佳实践、案例研究等问题开展有效对话。目前看来，主要的大国并未在双边层面开展相关的对话，由此带来的后果是相互之间的猜忌和不信任程度的增加。令人值得警惕的是，很多人把中国与美国在包括人工智能在内的高科技领域的竞争比作另一个“星球大战”，或者所谓的“科技冷战”，不仅导致了双方之间相互将对方视为敌手，甚至导致了对科学研究、供应链、产品的人为设限。因此，应通过相应的对话机制，通过有效的沟通来寻求更多的合作点，避免由于相互猜忌导致的恶性竞争。

总体而言，国家安全风险是人工智能发展过程中各国必须直面的挑战，在面临共同的威胁时，最优的做法是携手应对，通过相应的国际合作机制来降低安全威胁，增加合作空间，让人工智能更好地服务于人类社会的发展，增进人类社会福祉。

参考文献

[1]Executive Office of the President, Artificial Intelligence, Automation and the Economy[EB/OL].(2016-11-30).https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF.

[2] [英] 马丁·福特 . 机器人时代：技术、工作与经济的未来 [M]. 王吉美 , 牛筱萌译 . 北京：中信集团出版社，2015:235-238.

[3]Jochen Kruppa.Risk Estimation and Risk Prediction Using Machine-Learning Methods[J]. Human Genetics, 2012,131(10):1639-1654.

[4] 董青岭 . 机器学习与冲突预测——国际关系研究的一个跨学科视角 [J]. 世界经济与政治，2017(07):100-117.

[5]IEEE.Ethically Aligned Design[EB/OL].(2017-12-23).http://standards.ieee.org/news/2016/ethically_aligned_design.html.

[6] 中国电子技术标准化研究院 . 人工智能标准化白皮书（2018 年版）[EB/OL].(2018-01-24).http://www.cesi.ac.cn/201801/3545.html.

[7] 吴汉东 . 人工智能时代的制度安排与法律规则 [J]. 法律科学，2017,35(05):128-136.

作者介绍

封帅，上海国际问题研究院助理研究员，主要研究方向为人工智能和大国关系。

鲁传颖，上海国际问题研究院副研究员，主要研究方向为网络安全与网络空间治理。

National Security in the Era of Artificial Intelligence: Threats and Governance

FENG Shuai, LU Chuan-ying
(Shanghai Institutes for International Studies, Shanghai 200233 ,China)

[Abstract] Artificial intelligence is becoming the most important technical field affecting the future global society. Countries around the world have introduced guidance strategies to assist the innovation and development of artificial intelligence. However, as a subversive technology, Artificial intelligence contains a lot of risks and threats.The development of artificial intelligence technology will not only lead to problems in law and ethics, but also risks and threats in the field of national security. Based on the trend of artificial intelligence technology and applications, this article analyzes the security risks faced by countries in different fields, which not only helps to prepare for security, but also helps to deal with the relationship between security and development. Strengthening the effective response to risks can better guarantee the development of artificial intelligence technology. On this basis, the state can build a risk analysis and response framework from the aspects of raising risk awareness, improving governance system, strengthening supervision capability and exploring international cooperation, and improving the resilience of national security.

[Keywords] Artificial Intelligence;National Security;Governance