人工智能与国家安全：主要内涵及美国的战略认知

李 辉，王迎春
（上海市科学学研究所，上海 200235）

摘 要：本文分析了人工智能安全的主要内涵及对国家安全的意义，并基于美国政府的相关行动和重要智库发布的系列报告，讨论了美国应对人工智能安全问题，尤其是人工智能对美国国家安全的影响，所持有的态度和思路。

关键词：美国；人工智能；国家安全；治理
中图分类号：T18；G306 **文献标识码**：A DOI：10.3772/j.issn.1009-8623.2020.02.004

人工智能安全问题是当前人工智能发展中的关键问题，安全问题不能有效解决，人工智能产业将不可能健康发展。但是，由于人工智能应用的日益广泛，涉及的安全问题也越来越多元化。问题域的内容决定了政府治理的解决方向。本文梳理了当前涉及人工智能安全的主要问题，并分析了美国政府和智库从国家安全视角对人工智能安全问题的治理应对。

1 人工智能安全问题包含的主要内容

从目前的研究来看，人工智能安全问题主要分为两类，一是人工智能助力解决传统的安全问题；二是人工智能加剧或者带来了新的安全问题。前者体现了人工智能技术对传统信息安全行业的赋能，比如在网络攻击检测方面，安全人员可以利用智能算法对海量数据进行预处理，从而降低数据处理压力并作出正确判断。后者反映了人工智能应用影响的复杂性。当前，人们讨论人工智能安全问题主要指后者，本文讨论仅指后者。

人工智能加剧或者带来新的安全问题，主要表现在 3 个方面：人工智能技术自身不完善引起的安全问题；人工智能被不当使用（恶意使用或者一些无意不当使用）带来的安全问题；人工智能正当使用带来的安全问题。另外，从国家竞争角度，优先掌握新技术会带来国家竞争优势，落后国则限于被动，因此让自己国家处于人工智能发展的前沿也属于国家安全问题。当然各国争先恐后发展人工智能也可能带来无底线发展的伦理道德风险。

1.1 人工智能系统本身的不成熟性带来的风险

美国一份军事报告认为当前的人工智能（特别是深度学习）在 7 个方面（可靠性、可维护性、可问责性、可验证性、调试能力、脆弱性、可攻击性）表现得“能力薄弱”[1]。何积丰院士认为人工智能系统本身的技术安全包括数据安全、网络安全、算法安全、隐私安全 4 个方面。其中数据安全和网络安全是信息技术领域的传统议题，而算法安全则是指“如果算法正确，其结果便能够令人满意，而算法错误则会导致安全问题”；隐私安全主要指解决用户和相关者的隐私保护问题。目前讨论较多的是算法安全问题[2]。周鸿祎认为人工智能技术本身

第一作者简介：李辉（1981—），男，副研究员，主要研究方向为科技革命与产业变革及社会影响。
通讯作者简介：王迎春（1983—），男，副研究员，主要研究方向为创新变革与创新治理、科学技术与社会。邮箱：ycwang@stcsm.gov.cn
收稿日期：2020-01-25

的安全风险包括 4 方面：传感器可以被攻击，训练数据可以被污染，内部算法可以被欺骗，实现平台有漏洞 [3]。一些专家则从人工智能系统的 4 个关键环节——数据输入（传感器）、数据预处理、机器学习模型和输出，分析了相应的安全隐私风险 [4]。

1.2　人工智能系统的不当使用带来的安全风险

人工智能的技术能力，为黑客等恶意攻击方提供了新的技术手段。黑客利用人工智能技术，将使得传统的网络攻击智能化。人工智能的恶意应用是指利用人工智能技术进行攻击，以危害相关个人、组织或国家安全的行为。牛津大学的一份报告对此有专门的分析 [5]。一方面，黑客利用人工智能技术发动网络攻击，攻击规模和攻击效率将发生乘数效应。传统的网络攻击，攻击规模和攻击效率往往不能两全，而且需要大量的人力、物力和专业知识，而一旦通过人工智能系统攻击，成本将大大降低。攻击者可能并不需要太高的专业知识储备，并且会提升攻击的速度，增加可攻击的潜在目标。使用人工智能来执行攻击任务将不存在这样的问题，而且有可能扩大与劳动密集型网络攻击（如鱼叉式网络钓鱼）相关的威胁。另一方面，黑客利用人工智能技术发动物理攻击，可以完成传统网络攻击不可能完成的任务，如使用智能的无人机或其他系统攻击有关的目标。在人工智能时代，黑客还可以利用人工智能依赖数据的技术特点，通过提供假数据“数据下毒”，从而让人工智能系统做出错误的判断。“数据下毒”能够让很多智能系统变得混乱，成为人工智能时代的一种新的攻击手段。当然除了一些恶意使用，也存在无恶意的人工智能应用衍生出不良后果。

1.3　人工智能正当应用带来的经济社会安全问题

人工智能带来的产业革命，和历次产业革命一样，会对经济社会政治具有极强的重塑作用。在此过程中不可避免地带来安全问题。有学者列举了人工智能对经济、政治、军事、法律、思潮 5 个方面的负面影响 [6]。从目前的研究来看，比较显著的此类安全问题有代表性的 4 种。（1）人工智能取代就业岗位影响社会稳定。智库对人工智能影响劳动力市场的方式有很多预测。事实上人工智能对就业的影响讨论非常之多。虽然具体的数值仍存在很大争议（比如牛津大学两位学者的研究认为美国 47% 的职业都会被机器替代，引起了极大争议 [7]），但是如果人工智能的普遍应用在一个较短的时间内实现，对就业市场造成冲击并可能威胁社会稳定已成智库共识。（2）人工智能催生新旧动能转换引发政治冲突。人工智能的发展可能会引发一次新的产业革命，重塑世界经济中的利益格局、劳资关系。当前的一些政治冲突，与技术变革引发的社会矛盾都有一定关系。人工智能可能产生的不稳定性，也推动了全球民粹主义、民族主义运动的兴起。（3）人工智能制造假新闻。人工智能通过数据学习形成假图片、假声音、假视频等等，挑战“眼见为实”，造成舆论动乱，引发社会混乱。（4）人工智能放大极端思想、对公众进行大规模的精准误导。人工智能系统会收集、分析用户的行为数据，针对性地推送用户喜欢的内容，这种智能的过滤和筛选，容易导致个人陷入自己感兴趣的小圈子。这种应用也导致了新的社会风险。个性化推送新闻，容易引导舆论甚至误导舆论，尤其是一些假新闻引导的舆论，很容易产生政治风险。

1.4　国家间“无底线竞争”带来的安全问题

作为未来经济成功和战场决胜的关键因素，人工智能的领先地位对宏观的权力平衡和国际竞争而言至关重要。人工智能革命会引发全球各国力量、权力的变化，冲突性质的变化和军备竞赛。军备竞赛中存在潜在的风险，即为了胜利，不顾伦理法律的制约和技术上的不安全性。由于各国的人工智能军备竞赛往往都不公开透明，所以这类安全问题难以评估。另外，在军备竞赛中，世界各国的军队都在竞相制造更多的自主无人机、导弹和网络武器。更大的自主性意味着战场的反应速度更快。而人工智能武器泛滥的结果，就是可能会造成一些没有能力获得高精尖武器的国家或者实体获得远程精确打击能力。人工智能的发展将提供新型便捷化和商业化的武器，这也预示着可能会造成人工智能武器泛滥。如果不加控制地发展下去，战争的形式可能发生很大的变化，从而造成世界安全越发不可控的结果 [8]。

总体来说，前 3 类问题属于全人类共同面临的安全问题，而第 4 类属于国家安全问题。当然国家安全如果处理不当，也会给全人类带来安全问题。

2 从智库研究看美国对人工智能安全问题的认知

美国高度关注人工智能安全所涉及的问题。人工智能安全问题已经成为美国政府治理和智库研究的重要议题，尤其是人工智能与传统的国家安全领域相结合的新领域。

2.1 美国对人工智能安全的重点关注

2018 年 8 月，美国成立人工智能国家安全委员会（National Security Commission on Artificial Intelligence），旨在从国家安全角度对人工智能发展进行综合评估，并向美国国会和白宫提供政策建议。美国人工智能国家安全委员会计划在 2020 年最终报告中提出更全面的分析和建议[9]。

除了人工智能国家安全委员会，围绕人工智能与安全问题，美国还形成了体系化的治理机制。政府机构层面，美国国防部（United States Department of Defense）、白宫科技与政策办公室（The White House Office of Science and Technology Policy）、美中经济安全审查委员会（U.S.-China Economic and Security Review Commission）以及国会研究局（Congressional Research Service）等机构都将人工智能安全问题作为重要议题并发布系列官方报告。智库层面，贝尔弗科学与国际事务研究中心（Belfer Center for Science and International Affairs）、布鲁金斯学会（Brookings Institution）、伍德罗·威尔逊国际学者中心（Woodrow Wilson International Center for Scholars）、战略与国际研究中心（Center for Strategic and International Studies）、战争学院（Army War College）以及兰德公司（Rand Corporation）等纷纷发布相关研究报告[10]。尤其值得关注的是新美国国家安全中心，围绕美国所面临的人工智能安全问题进行了专题研究，并发布了系列报告。

智库的研究议题反映了美国政府的关注焦点。比较有影响力的有以下几份报告。

知名智库兰德公司密切关注人工智能在“安全性和就业”方面的影响，其发布的《人工智能对未来安全和就业的影响》指出，人工智能的应用普及将使数据隐私、消费者公平等问题日趋凸显；自动化和人工智能也将深刻影响就业。该报告认为人工智能安全的重点问题还包括：人与人工智能信息处理的差异；依赖人工智能增加了系统脆弱性；人工智能有可能造成经济和社会的快速破坏；人工智能具有地缘政治影响[11]。

美国国会研究局发布的《人工智能和国家安全》报告，关注军用人工智能议题，包括：如何平衡开发人工智能的商业投资和政府投资；军用人工智能开发的国防采购措施；在必要情况下对人工智能开发的监督；在人工智能和自主系统的研发与伦理方面如何取得平衡；为将人工智能应用融入军事领域必须变更的立法或法规；如何帮助美国在全球的人工智能竞争中取得成功[12]。

哈佛大学贝尔弗中心 2017 年发布《人工智能与国家安全》报告，认为人工智能的未来发展有潜力成为一种变革性的国家安全技术，与核武器、飞机、计算机和生物技术并驾齐驱，都导致美国国家安全共同体的战略、组织、优先事项和资源分配发生重大变化。人工智能的进步将通过推动军事优势、信息优势和经济优势 3 个方面的变化来影响国家安全。以“整个政府”为框架，报告提出了美国国家安全政策对人工智能技术的 3 个目标，并提出了 11 项建议[13]。

从这些报告的发布时间和内容来看，美国智库对人工智能安全问题的讨论有很强的前瞻性，也反映了美国各界尤其是政府关心的人工智能时代的国家安全问题。

2.2 新美国国家安全中心对人工智能与国家安全问题的系列研究

新美国国家安全中心是围绕人工智能安全的最为代表性的智库，他们发起了“人工智能和全球安全倡议”，专门成立了由美国前政府高级官员、私营企业领袖和学术专家组成的人工智能和国家安全工作组，探讨美国应如何应对人工智能所带来的国家安全挑战。该倡议的研究议程包括 5 方面：国际上的权力转移；冲突性质的变化（如人工智能的军事应用改变了战争性质）；危机的稳定性（如由人工智能系统的扩散引发的军备竞赛对危机稳定性的影响方式和影响程度）；人工智能技术安全性问题；国际合作[14]。

新美国国家安全中心围绕“人工智能和全球安全倡议”发布了一系列报告。根据内容可以看出其主要关心基本议题、未来国际关系、未来美国地位；

未来军事应用。

基本议题方面，如《人工智能：每个决策者都需要知道什么》，这份报告希望厘清人工智能和机器学习的可接受边界，核心的议题包括：确保美国在人工智能研究和创新领域的领导地位；授权联邦政府利用人工智能机会；确保在国家安全应用中安全可靠地使用人工智能；准备反击人工智能的恶意使用[15]。这份报告是美国讨论人工智能安全问题的基本核心议题。所以从国家的视角，在新美国国家安全中心看来，美国在人工智能时代最大的安全问题是确保技术的领先。

国家间竞争方面，如《人工智能时代的战略竞争》，该报告主要讨论美国在人工智能引发的变革中的国际地位。历史上的工业革命改变了力量平衡、国际竞争和国际冲突，该报告从工业革命的视角讨论人工智能的影响。第一次工业革命使欧洲和美国的生产率大幅度提高，首先是在英国，然后是其他国家。第二次工业革命导致了新一轮的国际竞争，19 世纪末 20 世纪初是一个多极化的环境，英国、法国、德国和俄罗斯在欧洲竞争，日本和美国也在崛起。专家们对是否发生了第三次工业革命、第三次工业革命的确切内容意见并不一致。然而，在 20 世纪 70 年代末和 80 年代初发生了明显的变化，当时微处理器、全球生产链和电子产品的结合推动了一股创新浪潮，创造了互联网、全球定位系统和一系列其他技术。这个时代在 20 世纪 90 年代初成熟，美国在这次浪潮中一家独大，因此，毫不奇怪美国在经济上领导着世界，与谷歌等巨头公司合作，在军事上领导着信息时代的武器。从本质上讲，第三次工业革命不是导致权力过渡或激烈的国际竞争，而是让美国一骑绝尘。历次工业革命的共同点，是战争性质和国家力量的转变。报告进而分析了人工智能对国际上权力的改变，可能包含以下有几种：拥有大量正确类型的数据；人工智能人才；计算资源；激励和协调有效采用人工智能的组织；公私合作；行动的意愿——对人工智能的监管、限制由于各国的伦理准则不同而有所差别[16]。在这篇报告中，新美国国家安全中心进一步分析了如何确保技术领先。

未来国际关系方面，如《人工智能与国际安全》，报告针对未来的国际关系进行了深入分析。提出了几个核心问题：我们最终会进入什么样的世界？人工智能是否开创了繁荣与和平的新时代？它是否会导致全球舞台上权力平衡的转变？人工智能会导致大规模的混乱和政治动荡、民族主义和保护主义的上升吗？人工智能是将控制信息的权力集中在少数人手中，还是会继续推进计算机、网络和社交媒体所带来的信息民主化？竞争信息的不和谐会导致从真相转向独裁主义和部落主义吗，还是民众的智慧会随着对真相和温和派政策的融合而胜出？人工智能带来的技术机遇塑造了未来，但并不决定未来。国家、团体和个人对于如何使用和响应人工智能的各种用途都有选择。他们的政策反应可以指导、限制或鼓励人工智能的某些应用。为了应对未来的挑战，美国将需要采取一项国家战略，以便在减轻破坏性影响的同时，如何利用人工智能的好处[17]。所以在这份报告中，在新美国国家安全中心看来，美国在人工智能时代的安全问题是，确保能妥善应对人工智能带来的政治、经济、社会等方方面面的风险。

军事方面，如《每秒一百万个错误》，世界各地的军队都在竞相制造更多的自主无人机、导弹和网络武器。更大的自主性使得战场上的反应更快，即将到来的“战场奇点”中，战斗的速度会使人类的决策能力黯然失色[18]。再如《技术轮盘赌：当许多军队追求技术优势时，管理失去控制》，这份报告认为引入人工智能这样复杂、不透明、新颖和交互式的技术，将产生不可预知的事故。因此美国国家安全机构可能将失去对美国创造的这些技术的控制[19]。在这两篇报告中，新美国国家安全中心着重强调了在军事竞争中，如何做到领先同时风险可控。

3　结语

人工智能的风险具有全球属性，其威胁超越国界。因此针对人工智能技术本身的不稳健、恶意应用以及对经济社会的复杂影响，全球需要齐心合力达成共识，共同解决。但是从国家竞争角度考虑，各国不可避免要争夺技术领导权和发展权，对人才的争取、技术的垄断等也是从国家安全视角考虑人工智能问题的一个重要方面。美国政府和智库的应对思考具有一定的前瞻性和参考价值。中国在人工智能的发展中，既需要考虑人工智能带来的共性安全问题，也需要从抢抓发展机遇的角度考虑发展和

安全问题。■

参考文献：

[1] Potember R. Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD[R]. McLean: The Mitre Corporation, 2017.

[2] 何积丰 . 安全可信人工智能 [J]. 信息安全与通信保密，2019（10）：5-8.

[3] 周鸿祎 . 人工智能安全及应对思考 [J]. 民主与科学，2019（6）：35-39.

[4] 陈宇飞，沈超，王骞，等 . 人工智能系统安全与隐私风险 [J]. 计算机研究与发展，2019，56(10)：2 135-2 150.

[5] Brundage M, Avin S, Clark J, et al. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation[R]. arXiv: 1802.07228, 2018.

[6] 封帅，鲁传颖 . 人工智能时代的国家安全：风险与治理 [J]. 信息安全与通信保密，2018（10）：30-49.

[7] Frey C B, Osborne M A. The future of employment: How susceptible are jobs to computerisation?[J]. Technological forecasting and social change, 2017(114): 254-280.

[8] Hunter A P, Sheppard L R. Artificial Intelligence and National Security: The Importance of the AI Ecosystem[R]. Washington D.C: Center for Strategic and International Studies, 2018.

[9] 高芳 . 美国发布人工智能与国家安全中期评估报告 [J]. 科技中国，2020（3）：95-97.

[10] 刘国柱，尹楠楠 . 美国国家安全认知的新视阈：人工智能与国家安全 [J]. 国际安全研究，2020，38（2）：135-155，160.

[11] Osoba O A, Welser W. The risks of artificial intelligence to security and the future of work[M]. Santa Monica: The RAND Corporation, 2017: 23.

[12] Sayler K M. Artificial Intelligence and National Security[R]. Washington, DC: Congressional Research Service, 2019.

[13] Allen G, Chan T. Artificial intelligence and national security[M]. Cambridge, MA: Belfer Center for Science and International Affairs, 2017: 36.

[14] CNAS. Artificial Intelligence and Global Security Initiative Research Agenda[R]. Washington, DC: Center for a New American Security, 2017.

[15] Scharre P, Horowitz M C. Artificial Intelligence: What Every Policymaker Needs to Know[R]. Washington, DC: Center for a New American Security, 2018.

[16] Horowitz M C, Allen G C, Kania E B, et al. Strategic competition in an era of artificial intelligence[R]. Washington, DC: Center for a New American Security, 2018.

[17] Horowitz M C, Allen G C, Saravalle E, et al. Artificial intelligence and international security[R]. Washington, DC: Center for a New American Security, 2018.

[18] Scharre P. A million mistakes a second[EB/OL]. (2018-09-12)[2020-01-13]. https://foreignpolicy.com/2018/09/12/a-million-mistakes-a-second-future-of-war/ .

[19] Danzig R. Technology roulette: Managing loss of control as many militaries pursue technological superiority[R]. Washington, DC: Center for a New American Security, 2018.

Artifical Intelligence and the National Security: Main Impacts and American Theory

LI Hui, WANG Ying-chun

(Shanghai Institute for Science of Science, Shanghai 200235)

Abstract: This paper analyzes the main connotation of AI security and its significance to national security, and based on the relevant actions of the US government and a series of reports issued by important think tanks, discusses the attitudes and ideas of the US to deal with the issue of AI security, especially the impact of AI on national security.

Key words: the U.S.; artificial intelligence; national security; governance