VIEW | 视野

美国智库探讨人工智能与国家安全的关系

■编译 / 高芳（中国科学技术信息研究所）

2017 年 7 月，美国权威智库哈佛大学贝尔佛科学与国际事务中心发布主题为《人工智能与国家安全》的报告，分析了人工智能技术通过变革军事、信息与经济优势将对国家安全产生的颠覆性影响，基于核、航空航天、网络和生物技术四种技术发展的经验教训，分别围绕确保 AI 技术领先、支持 AI 和平应用（军事）和商业应用、降低灾难性风险 3 大目标提出 11 项具体建议。

受美国情报高级研究计划署（IARPA）委托[1]，贝尔佛中心围绕人工智能与国家安全的关系开展专题研究。报告认为，人工智能对国家安全的影响至关重要，现有机器学习技术在提升劳动密集型工作自动化程度方面已具备相当的潜力，而近年来机器学习和 AI 的快速进步正成为战争走向自动化的历史转折点，虽然美国军方和情报机构正谋划在更广泛的领域应用人工智能技术，但实际上最具颠覆性的 AI 应用尚未出现。未来随着 AI 的持续进步甚至加速发展，AI 将像核、航空航天、网络和生物技术一样，有望成为影响国家安全的变革性技术。

一、人工智能技术对国家安全的变革性影响

1. 人工智能重塑军事优势。互联网、机器人和自主系统的广泛应用将增强非国家行为体[2]（Non-State Actors，NSA）和民族国家（Nation-State）实力。短期来看，人工智能的进步很可能会催生更多可直接参与战争的自主智能机器人，并加速有人作战模式向无人作战模式的转变。中

期来看，机器人和自主系统将逐渐拥有自然界本就拥有的多种能力，而长远来看这些能力将为军事领域和战争格局带来革命性变化。比如，致命自主武器成为军事主力，军事实力不再与人口规模、经济实力相匹配，无人系统集群作战技术改变作战模式，机器人暗杀成为寻常行动却难究真凶，移动机器人携带简易爆炸装置使恐怖分子获得低成本恐怖袭击能力，自主系统之间的非常规交互可能导致不可预测的后果，网络武器更加频繁的用于作战，在军事系统中应用机器学习产生新型漏洞并催生新型网络攻击手段，人工智能军事系统一旦被盗或者非法复制将使AI网络武器被恶意使用。

2. 人工智能重塑信息优势。人工智能将大幅提升数据收集和分析能力、文本和多媒体数据的生成能力，而正是这种数据生成能力将对未来的政治宣传、战略欺诈以及社会工程等产生重大影响。拥有先进人工智能分析系统的国家将拥有战略决策的决定性优势，主张独裁和专权的政治宣传将越来越难以与现实情况相区分，全方位无死角的超级监视将游击战和叛乱活动等扼杀在萌芽中。AI伪造信息广泛存在将侵蚀社会信任体系，虚假新闻有可能变得更加令人信服，伪造信息与网络攻击、社交媒体机器人网络相结合将严重威胁政治经济稳定。在军事情报领域，伪情报制作变得更加便利，当AI被用来伪造国防部指令和政治声明并在互联网上大肆传播，或用来模仿军事或情报官员下令分享敏感信息或要求采取行动时，局面将变得异常严峻。

3. 人工智能重塑经济优势。人工智能是科技创新的超级引擎，基于AI推动科学实验过程自动化、从上千篇文献中挖掘出新知识、自动生成并优化工程设计等将加速科学发现与技术发明过程，谁占据了AI技术的研发优势谁就具备了技术与经济优势的自我强化能力。谁开创了AI创新应用的先河，谁就有望成为下一代创新的先行者，进而赢得经济、军事等领域长久的战略优势。不断进步的自动化将使失业问题更加严峻，未来若真的出现大规模失业现象，那么技术先进国家有可能面临“资源诅咒[3]”问题——“自动化”作为一种“资源”，其对一国的经济增长并不构成充分的有利条件，反而是一种限制。AI的经济影响成为一种新型“武器”，美国国家安全以及美国领导的联盟体系将面临前所未有的挑战。

二、发展人工智能需要借鉴的经验教训

分别从破坏性潜能、成本、对智力资源的需求程度、军民两用潜能以及监管难度共5个维度对核、航空航天、网络和生物技术四种技术进行评估，同时梳理了美国政府对每种技术的管理模式，以确保美国的技术领先优势、确保民用/军事领域同时获益、有效实施灾难性风险管控为坐标分别评价了四种技术管理模式的有效性，厘清了应对各种挑战的具体措施。仅从技术管理角度，与上述四种技术相比人工智能有可能引发最糟糕场景：

一是AI具有最高级别的破坏性潜力，未来通用人工智能（General AI）和超级智能系统对人类存在构成威胁。

二是AI研发成本具有多样性，众多志在研发AI前沿技术的公司会投入数十亿美元的研发经费，而使用开源平台进行产品开发的小公司则无需投入太多。

三是AI对专业人才的需求同样具有多样性，与一般性AI应用不需要太多顶尖人才不同，进行前沿探索的顶尖人才总体不足，能够将AI商用技术应用在军用系统中的高端人才也极度短缺。

四是AI是典型的军民两用技术，商业和军事领域对AI硬件基础设施和人才等方面的需求是相同的，而某些军事应用（比如自主武器）中更需

要一些非 AI 领域的专家才能将 AI 的潜能充分激发出来。

五是 AI 具有最高级别的监管难度，其原因在于 AI 军民两用技术的天然属性使得难以辨别哪些 AI 活动具有潜在的破坏性，要对自主武器系统进行最直接的监管基本不具备可操作性，同时实际参与高级 AI 系统研发部署的人数已远远超过核与航空航天领域。

类比以上四种技术的发展历程，总结出 AI 发展需要借鉴的经验教训：一是“激进”的技术变革有可能催生全新的政府管理理念，随着 AI 影响越来越广泛而明显，当前一些“激进”的应对措施和做法等将被证明是有效的。二是军备竞赛不可避免但终归还是可以管控，尽管不可能完全禁止人工智能在军事领域、在国家安全领域的应用，尽管在 AI 技术尚未成熟之前仍有诸多未知因素，仍然要积极探索实现安全有效的技术风险管控目标。三是政府对商用 AI 既要鼓励又要适当限制，鼓励的目的是促进经济增长，然而“稍不留神”有可能损害国家安全。四是将“确保 AI 安全可靠”落到实处，具体就是要设立专门机构（提供经费、人力成本等多种资源保障，树立官方的权威性）负责监管整个政府和商用 AI 领域的安全问题。五是国家战略有可能伴随技术的发展而变化，政府当务之急是要塑造 AI 在军事和情报领域应用的技术优势。

三、发展建议

为确保美国人工智能的领先地位，建议美国国防部（DoD）开展聚焦 AI 的军事演习以甄别潜在的颠覆性军事创新，资助以人工智能技术及其影响为主题的各种长期战略分析与评估，优先支持能带来稳定收益和降低关键风险的 AI 研发；美国国防和情报机构应该重点投资“反人工智能”的进攻和防御能力。

为确保和平运用人工智能技术，建议美国国防高级研究计划局、美国情报先进研究计划署（IARPA）、海军研究署和国家科学基金会应增加与 AI 相关的基础研究的经费支持；国防部应发布关于军民两用 AI 能力的信息需求（RFI）；In-Q-Tel 公司[4]应为促进国家安全部门与商用 AI 公司间的合作提供额外的资源。

为实现有效的 AI 灾难性风险管理，美国国家安全委员会、国防部与国务院需研究 AI 应用对美国的潜在影响，并依此制定相应的限制条款；国防部和情报部门应当建立专门的 AI 安全机构；DARPA 应对 AI 系统的失效保护及安全使用提供研究经费支持；美国国家标准技术研究所（NIST）和美国国家安全局（NSA）应探索应对人工智能造假的方案。科技

注：

1 IARPA, Intelligence Advanced Research Projects Activity 成立于 2006 年，是美国政府以 DARPA 为蓝本设立的情报科技创新机构，致力于投资高风险、高收益的研究计划，过去 10 年来 IARPA 已成为量子及超导计算领域学术研究的最大资助方，并在机器学习、语音识别、图像分析、人脸识别、自动视频分析等领域有大量的资金投入。

2 NSA 指影响力巨大却不属于任何国家的群体。

3 资源诅咒是一个经济学的理论，是指从长期的增长状况来看，那些自然资源丰裕、经济中资源性产品占据主导地位的国家反而要比那些资源贫乏国家的增长要低许多；尽管资源丰裕国家可能会由于资源品价格的上涨而实现短期的经济增长，但最终又会陷入停滞状态，丰裕的自然资源最终成为“赢者的诅咒”。

4 In-Q-Tel, 一家非营利性的高科技风险投资公司，以增进其所投公司与国家安全机构之间的联系，确保美国情报机构拥有最先进的技术和最优的情报能力。