☞军事智能☜

人工智能防御技术研究与启示*

赵擎天,李立伟,陈鑫,侯立志

(军事科学院 系统工程研究院,北京 100101)

摘要:人工智能技术在军事领域的广泛应用,催生战争形态不断演进。为应对智能化战争带来的挑战,通过分析人工智能技术发展趋势及其地位作用,讨论了在军事领域开展人工智能防御技术研究的必要性;从“硬打击”和“软打击”2个角度,给出了防御反制人工智能武器系统的技术方法和AI安全防御架构;重点分析了污染打击、逆向打击、后门打击和对抗打击的技术原理;最后,从顶层设计、应用审查、数据治理、基础研究、指挥艺术和人机混合智能6个方面给出了加强人工智能防御技术研究的措施建议。

关键词:人工智能;防御技术;智能化战争;机器学习;硬打击;软打击

DOI:10.3969/j. issn. 1009-086x. 2024. 03. 007

中图分类号:E919;TP18;TJ0 **文献标志码**:A **文章编号**:1009-086X(2024)-03-0055-09

引用格式:赵擎天, 李立伟, 陈鑫, 等.人工智能防御技术研究与启示[J].现代防御技术,2024,52(3):55-63.

Reference format: ZHAO Qingtian, LI Liwei, CHEN Xin, et al. Research and Enlightenment of Artificial Intelligence Defense Technology[J].Modern Defence Technology,2024,52(3):55-63.

Research and Enlightenment of Artificial Intelligence Defense Technology

ZHAO Qingtian, LI Liwei, CHEN Xin, HOU Lizhi

(Institute of Systems Engineering, Academy of Military Sciences, Beijing 100101, China)

Abstract: The wide application of artificial intelligence technology in the military field has led to the continuous evolution of war forms. To meet the challenge of intelligent warfare, this paper analyzes the development trend and the role of artificial intelligence technology and discusses the necessity of developing artificial intelligence defense technology in the military. From the view of "hard strike" and “soft strike”, the technical ideas of defense against anti-AI weapon systems and AI security defense architecture are given. The technical principles of pollution strike, reverse strike, backdoor strike, and confrontation strike are analyzed emphatically. Suggestions for strengthening AI defense technology research are given from six aspects: top-level design, application review, data governance, primary analysis, command art, and human-machine hybrid intelligence.

Keywords: artificial intelligence; defense technology; intelligent warfare; machine learning; hard strike; soft strike

* **收稿日期**:2023-03-29;**修回日期**:2023-06-12

第一作者简介:赵擎天(1986-),男,山东高唐人。博士生,研究方向为智能化装备体系设计。

E-mail: zhaoqingtian1986@163.com

0 引言

在大国竞争日趋激烈和世界多极化的背景下，人工智能作为一种力量倍增器，在推进军事领域技术创新方面是最活跃的因素，正在发挥着重塑战争规则的变革性作用。目前，人工智能已广泛应用于军事情报数据分析、指挥控制系统决策、军事网络链路设计、自主武器系统和无人作战平台等方面[1]。近年来，为了应对未来战场环境的复杂性，美国国防高级研究计划局 DARPA（defense advanced research projects agency）以人工智能技术为基础，推出了“马赛克战”概念，其具有低成本、模块化、分散灵活、自适应动态组合和跨域协同的特点。这表明未来战争将越来越多地应用科学计算、数据分析和智能算法等人工智能技术以驱动高度变化、不可预知的战场环境，从而促进战争进程不断加快。

人工智能系统的核心是机器学习算法，但机器学习算法具有统计和黑盒特性，这就导致人工智能系统容易被学习数据干扰，且具有不可解释性，使用人员又很难察觉系统运行过程中的问题，造成了固有的脆弱性[2]。因此，人工智能在军事领域中的广泛应用，必将催生基于人工智能固有弱点的防御反制人工智能武器系统的新技术和新方法，本文将这些新技术和新方法统称为人工智能防御技术。目前，人工智能防御技术主要包括“硬打击”和“软打击”2个方面。“硬打击”主要以传统的物理摧毁手段为主；“软打击”则是以人工智能算法的技术缺陷为突破口，采取数据欺骗和算法打击等手段对人工智能武器系统进行防御反制。

1 人工智能防御技术研究必要性分析

1.1 军事科技创新发展规律的内在要求

纵观世界军事科技发展历史，一种新技术应用于军事领域后，与其对抗的反制技术必将随之出现。例如，世界上第 1 辆坦克诞生于 1916 年的英国，而在 1917 年德国便发明了专门用于打击坦克的“K”型子弹，对于防空、反潜和反导技术的发展也遵循同样的规律。人工智能作为一种尖端技术，已经在世界军事技术领域引起了重大变革，这种变革将是颠覆性的，甚至可能从根本上改变战争形态和作战方式，这更加需要始终保持思想的清醒与思维的清晰，从新生事物上把握与发掘颠覆性影响因素，从渐进式演变中把握革命性变化内因，超前布局，打好主动仗。设计武器装备就是设计未来战争，研究发展人工智能防御技术就是顺应军事科技创新发展需求的必然选择。所以，加快发展人工智能防御技术的研究，既符合当前军事斗争准备的需要，也与军事科技创新发展的辩证统一规律相契合。

1.2 国际军事实力竞争的迫切需要

俄罗斯总统普京曾直言“谁能成为人工智能领域的领先者，谁就能统治整个世界”。由于人工智能技术具有显著的泛在性、赋能性和不确定性，世界各国已经争先恐后地踏上了人工智能竞赛的赛道。美国在人工智能领域居于全球霸主地位，以 DARPA 为代表的政府机构持续推动人工智能发展与应用，并且已经在军事领域进行了深入应用探索。俄罗斯于 2019 年 10 月批准了《2030 年前俄罗斯国家人工智能发展战略》，旨在促进俄罗斯在人工智能领域的快速发展和获得技术竞争力。英国一直是全世界研究人工智能的学术重镇，形成了以伦敦、剑桥、爱丁堡等高校集中城市为中心的人工智能产业集群。欧盟各国以及日本、韩国、新加坡和印度等国相继出台了人工智能国家发展战略，以期举全国之力在人工智能领域获得技术竞争优势。我国经过多年持续积累，在人工智能领域部分核心技术方面实现了重要突破。但不可否认的是，我国人工智能在基础研究、芯片、人才等方面同世界其他先进国家相比还有差距[3]。鉴于人工智能在战略竞争中的作用，在未来军事对抗中如何有效遏制人工智能武器系统，以取得竞争优势，是开展人工智能防御技术研究的出发点和落脚点。

1.3 打赢未来智能化战争的重要支撑

智能化战争核心是算法的竞争，人工智能防御技术的本质是利用智能对抗智能。目前，美军关于人工智能防御技术的概念公开资料较少，但美国国防部制定的多项计划已经透露出其通过技术霸权达到防御反制人工智武器系统的企图。2022 年 5 月 31 日，美国更新《国防部反无人机系统报告》，指出美军将大力发展反无人机武器及相关技术，企图在

反无人机领域获取绝对优势。美军在反无人机问题上不断加大投入，目的是通过在反无人机领域率先形成人工智能防御反制能力，进而巩固未来的竞争性战略优势，维护其全球霸权。在未来智能化战争中，人工智能技术极大增强了人对战场的认知能力，而如果想赢得战争，就必须在强化自身能力的前提下，采取各种措施削弱或消除对方认知能力，以使战场单向透明形成有利态势，而人工智能防御技术的研究将为建立这种优势提供支撑。

2 人工智能防御主要技术手段及AI安全防御架构

目前，人工智能技术的发展仍处于初级阶段，要实现真正的“通用智能”或“强智能”还需在基础理论研究上有重大突破[4]。人工智能技术在军事领域的应用深度和广度不断创新突破，但由于其存在固有技术缺陷，探索发展人工智能防御技术已经成为可能。

2.1 硬打击

“硬打击”主要指采取传统作战方式，如硬杀伤、捕获、摧毁或电磁破坏等方式，使敌人工智能武器系统丧失继续作战的功能和能力。目前，从防御反制策略分析，如图1所示主要有以下3种方式。

图1 “硬打击”主要方式

Fig. 1 Main “hard strike” modes

（1）利用电子战手段干扰或破坏人工智能武器系统的通信、导航、感知等功能，降低其效能或使其失控。电子对抗方式作用范围大、受天候影响小，可实施灵巧式干扰攻击。例如，在防御反制地面无人系统时，可以利用电子战手段阻断其遥控遥测链路、干扰其导航定位和环境感知能力，从而达到防御反制人工智能武器系统的目的[5]。

（2）通过动能、化学能和定向能等武器系统能量的释放达到毁伤目标物理特性的技术。其中，以投射动能弹丸和火药为主的火力打击方式，毁伤效果直接；以激光、微波和粒子束等武器为主的定向能武器，相比传统火力打击具有射束快、精度高、反应灵活和附带损失少等特点，是当前防御反制无人机等小型装备的有效手段[6]。

（3）利用自身的人工智能武器系统，如无人机、机器人等，来对抗或摧毁敌方的人工智能武器系统。以防御反制无人机为例，无人机群可以通过精准的协同配合作战摧毁敌方无人机，并且这种方式具有低成本、高效率和零伤亡等优势。随着人工智能技术的发展，以无人机为代表的人工智能武器系统之间的对抗将成为智能化战争的主要形态（表1）。

表1 “硬打击”主要方法分析

Table1 Analysis of the main methods of “hard strike”

方法	主要内容	存在问题
电子对抗	用电子战手段干扰或破坏人工智能武器系统的通信、导航、感知等功能，降低其效能或使其失控	
能量毁伤	通过动能、化学能和定向能等武器系统能量的释放达到毁伤目标物理特性的技术	看不见 杀不尽 打不起
智能对抗	利用自身的人工智能武器系统，来对抗或摧毁敌方的人工智能武器系统	

随着人工智能武器向隐身化、综合化、微型化、集群化、自主化和低成本等方向发展，“硬打击”这种防御反制人工智能武器系统的技术“看不见”、“杀不尽”和“打不起”的问题将越来越突出[7]。例如，水下无人潜航器UUV（unmanned underwater vehicle）目标特征小、隐蔽性强、机动灵活，依托复杂水下作战环境的有效掩护，仅靠现有探测预警手段难以及时发现目标[8]。在俄乌冲突中，俄罗斯投

入战场使用的“见证者-136”自杀式无人机造价仅2万美元，而乌克兰能对抗这种无人机的“毒刺”防空导弹是15万美元，德国援助的“IRIS-T SLM”地对空导弹价格则高达38万美元，若采用以上方式对抗俄罗斯无人机，乌克兰确实存在“打不起”的问题。

2.2 软打击

“软打击”主要指针对人工智能系统算法的固有弱点，通过实施数据欺骗、无规则动作或制造迷惑性战场态势，使敌人工智能武器系统判断错误、算法失效[9]。“软打击”技术的早期阶段，主要聚焦于训练数据污染和模型仿制重建上，但由于模型训练的机密性，一般很难做到。自2015年生成式对抗网络GAN(generating adversarial network)技术出现后，基于生成对抗网络的人工智能防御技术逐渐成为研究重点，以此为基础的各类打击和防御算法不断涌现。如图2所示，“软打击”主要有以下4种方式。

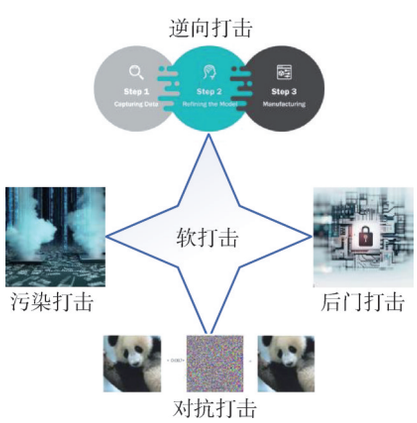

图2 “软打击”主要方式

Fig. 2 Main “soft strike” modes

2.2.1 污染打击

污染打击，又称数据投毒，指在人工智能系统模型创建过程中，将伪装数据、欺骗样本等加入到其训练和测试数据中，以破坏其学习数据的真实性，进而导致模型算法判断或决策出现偏差。例如，在战前通过对参战车辆进行个性化伪装，敌无人侦察机如果仍采用原始数据训练的AI模型，就很难识别伪装后的车辆，导致其侦察能力大大削弱[10]。所以在战时应加强对武器装备或军事设施的伪装防护，从而降低敌智能侦察系统的目标识别能力。

2.2.2 逆向打击

逆向打击是指对捕获的敌方的AI系统，通过将特定信息输入到模型并观察其输出响应的方式，达到模型重建的目的。逆向技术在工业领域是指对一项目标产品进行逆向分析及研究，以制作出功能相近，但又不完全一样的产品，其目的是在不能获得必要的信息的情况下，直接从成品分析推导出产品的设计原理。尽管机器学习引擎的内部机制神秘莫测，但研究人员已经发现这些黑箱的内容物可以被逆向，甚至可以完全复制。2016年8月，康乃尔科技学院、瑞士洛桑理工学院、北卡罗莱纳大学的计算机科学家，通过发问和分析响应，用目标AI的输出来训练他们自己的AI，使被克隆的AI预测准确率接近100%[11]。

2.2.3 后门打击

后门打击是指将有漏洞的AI模型部署在军事系统中，在战时触发该漏洞，从而使AI系统失效，或者控制AI系统给敌人造成更大威胁。实际工作中，程序开发人员往往只关心样本数据和模型参数，对底层框架的安全性考虑较少，这就为打击者留下了可乘之机。例如，打击者通过在神经网络模型中植入特定的带有后门的模型，使得AI系统对正常输入能够正确判断，但对特殊输入的识别会受打击者控制，而AI设计者或使用人员很难通过分析模型发觉后门的存在。这也提醒我们在开发利用人工智能技术的同时，应加强安全评估，加大基础理论研究，将核心技术牢牢掌握在自己手中。

2.2.4 对抗打击

对抗打击就是通过输入生成的对抗样本，迫使被打击的AI模型做出错误的响应。对抗样本是指在原数据集中通过人工添加不容易被发现的细微扰动所形成的样本，这类样本会导致训练好的模型以高置信度给出与原样本不同的分类输出。2015年对抗网络的创始人Ian Goodfellow提出了对抗打击的概念，他在一张熊猫的图片加入人为设计的微小噪声之后，人眼对扰动前后两张图片基本看不出区别，但人工智能模型会以99.3%的概率将其错判为长臂猿[12]。2019年8月，莫斯科国立大学、华为莫斯科研究中心公布了一项研究成果，只需将一张带有对抗

图案的纸条贴在额头上，就能让性能优越的人脸识别精度大大降低[13]。根据打击者的知识掌握程度可将打击分为白盒、黑盒和灰盒打击。白盒打击需掌握打击目标模型的完整知识，黑盒打击则完全不用了解目标模型的神经网络架构、参数等信息，而是通过模型的查询访问结果来生产对抗样本；灰盒打击需要了解目标模型的结构，但不必知道目标模型的具体参数[14]。由于对抗打击具有隐蔽性高、侵害面广、打击力大和迁移性强等特点，其作为一种人工智能防御技术，在军事对抗领域具有较大应用前景。“软打击”主要方法分析如表2所示。

表2 “软打击”主要方法分析

Table 2 Analysis of the main methods of "soft strike"

方法	主要特点	存在问题
污染打击	伪装数据、欺骗样本以破坏学习数据的真实性，进而导致模型算法判断或决策出现偏差	数据获取难 技术难度高
逆向打击	对捕获敌方的AI系统通过逆向分析，达到模型重建目的	
后门打击	在AI模型中嵌入病毒，战时触发该漏洞，从而使AI系统失效	
对抗打击	输入生成的对抗样本，迫使被打击的AI模型做出错误的响应	

2.3 AI安全防御架构

针对上文提到的人工智能“软打击”的各种方式，如何进行防御反制已经成为军事智能领域研究的重点。本文针对未来军事智能化发展需求，结合当前AI成熟对抗技术，设计开发了AI安全防御架构。本架构从预先防范、自身免疫和冗余设计等3个层次构建AI安全防御体系，以保证在战场激烈对抗环境下AI系统的有效性。如图3所示，AI系统的安全防御架构主要包括3部分：第1层是预先防范机制，该机制主要在数据采集、模型训练和模型部署的过程中，分别针对污染打击、后门打击、逆向打击和对抗打击等“软打击”方式进行针对性防御，例如通过训练数据过滤[15]、输入重构[16]和对抗样本检测等技术可以实现对污染打击和对抗打击的防范；利用模型剪枝[17]和差分隐私[18]技术可以对后门打击和逆向打击进行防御。第2层是自身免疫机制，主要是探索建立可解释人工智能系统，以数据可解释为基础，积极构建可解释和可验证的AI模型，通过AI模型的可解释性明晰系统输入、输出和过程数据的逻辑关系，从而实现对恶意攻击的甄别、清除和修复，提高AI系统的鲁棒性，实现自身免疫。第3层是冗余设计机制，该机制主要在AI系统部署时建立实施。当AI系统因攻击性能降低或功能失效时，根据预设的判断规则自动触发冗余保护机制，自动关闭被攻击的系统，同时启动备份系统或转入人工控制，从而最大限度降低系统风险。

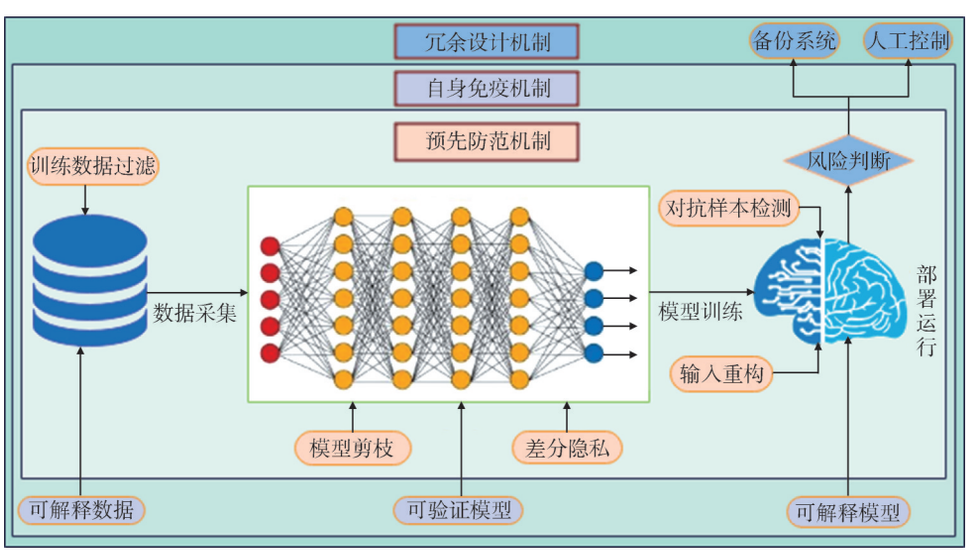

图3 AI安全防御架构

Fig. 3 AI security defense architecture

3 人工智能防御技术研究启示

面对复杂多变的国际发展环境和激烈对抗的严峻挑战，必须坚持实施自立自强的科技创新发展战略，加强人工智能防御技术机理研究，充分把握智能化战争制胜规律，不断提高科技创新对战斗力增长的贡献率。关于加强人工智能防御技术研究，主要有以下研究启示：

3.1 科学设计人工智能防御体系

坚持运用体系思维，将人工智能防御技术研究纳入智能化作战体系，统筹设计、一体推进。体系对抗是现代战争的重要特征，军事创新技术必须融入作战体系才能发挥最大作用。开展人工智能防御技术研究，必须把作战体系需求作为基点，科学设计人工智能防御体系，重点围绕作战运用、能力功能、系统支撑和应用装备等进行科学论证，综合确定最优的人工智能防御体系发展路线图。要充分尊重科学技术的发展规律，遵循先易后难、循序渐进的原则，按照模型、系统、体系发展路径，加强概念演示验证，以工程实践推动技术的落地见效。

3.2 建立人工智能应用审查评估制度

从竞争对抗角度考虑，降低人工智能技术风险和防御人工智能打击的有效手段，就是科学合理使用人工智能技术。2019 年 8 月，美国贝尔弗科学与国际事务研究中心马库斯·康米特发表《Attacking Artificial Intelligence：AI’s Security Vulnerability and What Policymakers Can Do About It》，专门分析了人工智能固有的数据依赖和“黑盒”问题带来的潜在安全威胁，并提出了建立人工智能安全合规计划的应对举措[19]。如图 4 所示，军事效能显著、安全性能好、使用成本低的人工智能系统应推荐使用；相反，军事效能一般，安全评估存在较大风险，使用成本较高的人工智能系统应坚决制止，从而避免重技术、轻效能的现象发生。因此，在军事领域应加强人工智能技术应用可行性分析，建立人工智能应用审查评估制度，综合考虑军事效能、安全风险和使用成本等因素，科学确定应用方式和使用程度。

3.3 提高军事数据治理能力

人工智能的作用机理就是在大量样本中发现其某一方面特征的规律性，进而做出确定性预测。人工智能防御技术的核心是军事数据，军事数据是人工智能武器系统作用发挥的前提，加强军事数据治理对增强数据可信度、提高决策科学化和人工智能武器系统防御能力具有重要作用[20]。因此在未来智能化战争中，关于军事数据的博弈是人工智能技术对抗的焦点，谁能掌控数据谁就能获得战争的主动权。在军事实践活动中，应在充分理解人工智能作用机理的基础上，强化数据保护意识，树牢保护数据就是智能防御的意识，不断提高军事数据治理和应对未来智能战争的能力。

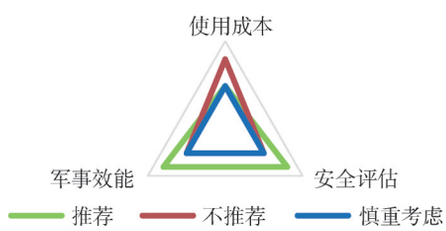

图 4 AI 系统应用审查分析模型

Fig. 4 AI system application review analysis model

3.4 加强人工智能防御技术基础理论研究

以深度学习为代表的人工智能技术在军事领域已初露锋芒，但由于军事数据存在积累样本少、数据缺乏标注和无法验证等问题，以无监督学习和小样本学习为基础的可解释的通用人工智能更适合军队需要。目前，可解释的通用智能基础理论还比较薄弱，需要加强研究、创新突破。在人工智能防御技术研究领域，应前瞻布局脑科学、计算机科学、哲学等跨学科基础研究，探索人工智能防御技术发展的客观规律和根本目标；在技术应用领域，在“硬攻击”方面，应加强多种探测技术复合的探测预警、定向能毁伤和智能对抗等武器系统开发研究，建立能有效应对不同维度、不同作战域人工智能武器的“硬打击”对抗体系。在“软打击”方面，应加强基于生成对抗网络的人工智能防御技术研发，同步推进自然语言处理、认知推理和训练技术研究，积极探索机器学习以外的革新技术路径。同时，对于如何评估人工智能系统的鲁棒性和增强系统对抗打击的能力，仍是一个基础性难题，需要坚持不懈持续攻关。

3.5 高度重视战争科学和指挥理论研究

战争科学是一门复杂性科学，战争复杂系统的

反身性告诉我们，一个战争理论从证明有效开始就逐步走向消亡的道路[21]。从这一理论出发，人工智能系统计算得越准确，反身性影响就越大，预测结果就越有可能被敌人利用。人工智能防御技术的本质是应对被感知的确定性和规律性，而“人”作为战争指挥艺术的创造者，可以通过先验知识和对战场环境变化的把握创造不确定性，而这种不确定性是现有人工智能无法计算预测的。因此在未来智能战争中，“人”的战争指挥艺术是防御反制人工智能的终极武器。在可预见的未来，人工智能仍旧无法完全取代人类在战争中的作用，无论人工智能如何先进，其承载的仍然是人的作战意图，智能战争的背后依然是作战方法、指挥艺术与意志品质的较量，未来战争的主动权仍旧牢牢掌握在人类手里。

3.6 探索军事人机混合智能发展路径

未来战场环境必将是人、机、环相互作用、高度融合的，战争形态也不仅仅是智能化的，更是智慧化的。在人工智能对抗防御方面，既有机器智能的对抗，也有人类智慧的博弈。机器智能在搜索、计算、存储和优化等方面优于人类智能，而人类在感知、推理、归纳和学习等方面具有机器智能无法对抗的优势。在应对“硬打击”和“软打击”时，可以将人的业务常识与经验引入到机器智能攻防决策的生成回路中，利用人对模糊、不确定问题分析与响应的优势，形成一种“1+1＞2”的增强型智能形态，也就是人机混合智能。探索以人机混合智能为基础的军事智能发展路径，提升指挥员主动介入决策问题求解过程的智能交流能力，突破人类和机器智能的双向流动瓶颈，建立机器和人类相互启发、互相学习、共同提高的机制，形成不断提升系统整体智能水平的持续反馈回路，使人工智能真正成为未来指挥作战的有力助手。

4 结束语

人工智能作为第4次工业革命的核心驱动力，已成为世界各军事强国激烈争夺的战略制高点。本文从人工智能防御技术在军事领域应用的必要性出发，分析给出了在军事领域开展防御反制人工智能武器系统的主要方法，最后通过综合分析给出了在军事领域开展人工智能防御技术研究的建议思考。当前，以ChatGPT为代表的AIGC（AI generated content）人工智能技术正掀起新一轮的人工智能革命，其在人机交互技术方面的重大创新为AI的迅速普及发挥了重要推动作用。展望未来，必须时刻紧盯AI前沿技术发展动态，加强人工智能技术机理研究，力争在人工智能防御技术赛道上取得竞争优势。

参考文献：

[1] 丹尼尔·阿瑞亚，梅格·金，李欣来. 人工智能对军事防御与安全的影响（译文）[J]. 信息安全与通信保密，2022(6)：67-73.
ARAYA D，KING M，LI Xinlai. The Impact of Artificial Intelligence on Military Defence and Security [J]. Information Security and Communications Privacy，2022(6)：67-73.

[2] 中国信息通信研究院. 人工智能白皮书（2022年）[EB/OL]. [2023-03-28]. https://www.scdsjzx.cn/scdsjzx/ziliaoxiazai/2022/4/18/d03a2d33b67d4c398ddfca504cf410ab/files/43b00b8feccd423ea2e2a4014e9d672a.pdf.
China Academy of Information and Communications Technology. Artificial Intelligence White Paper (2022) [EB/OL]. [2023-03-28]. https://www.scdsjzx.cn/scdsjzx/ziliaoxiazai/2022/4/18/d03a2d33b67d4c398ddfca504cf410ab/files/43b00b8feccd423ea2e2a4014e9d672a.pdf.

[3] 李睿深，石晓军，郝英好. 人工智能[M]. 北京：国防工业出版社，2022：32-37.
LI Ruishen，SHI Xiaojun，HAO Yinghao. Artificial Intelligence [M]. Beijing：National Defense Industry Press，2022：32-37.

[4] 刘伟. 人机融合：超越人工智能[M]. 北京：清华大学出版社，2021：10-14.
LIU Wei. Human-Machine Fusion：Beyond Artificial Intelligence [M]. Beijing：Tsinghua University Press，2021：10-14.

[5] 王伟，王钦钊，刘钢锋，等. 地面无人系统反制关键技术分析与综述[J]. 航空学报，2022，43(7)：16-40.
WANG Wei，WANG Qinzhao，LIU Gangfeng，et al. Countering Unmanned Ground System：A Review of Key Technologies[J]. Acta Aeronautica et Astronautica Sinica，2022，43(7)：16-40.

[6] 刘文学，王涛，李赵健伟，等. 反无人机装备发展现

状及趋势[C]//2022年无人系统高峰论坛(USS2022)论文集. 中国,西安: 中国工程院, 西北工业大学, 中国航天科工集团有限公司, 国防科技大学, 2022: 37-44.
LIU Wenxue, WANG Tao, LI Zhao jianwei, et al. Development Status and Trend of Anti-UAV Equipment [C]//2022 Unmanned Systems Summit Forum. Xi'an, China: Chinese Academy of Engineering, Northwestern Polytechnical University, China Aerospace Science and Industry Corporation Limited, National University of Defense Technology, 2022: 37-44.

[7] 张冬冬, 王春平, 付强. 国外无人机蜂群发展状况及反蜂群策略研究[J]. 飞航导弹, 2021(6): 56-62.
ZHANG Dongdong, WANG Chunping, FU Qiang. Research on the Development of UAV Swarm and Anti-bee Swarm Strategy Abroad[J]. Aerodynamic Missile Journal, 2021(6): 56-62.

[8] 迟蛟龙, 李洪权, 李新, 等. 近海反UUV战法研究[C]//2022年无人系统高峰论坛(USS2022)论文集. 中国,西安: 中国工程院, 西北工业大学, 中国航天科工集团有限公司, 国防科技大学, 2022: 69-73.
CHI Jiaolong, LI Hongquan, LI Xin, et al. Analysis of Counter-UUV Warfare in Offshore Waters[C]//2022 Unmanned Systems Summit Forum. Xi'an, China: Chinese Academy of Engineering, Northwestern Polytechnical University, China Aerospace Science and Industry Corporation Limited, National University of Defense Technology, 2022: 69-73.

[9] 韦正现, 王桐. 反智能化作战的研究启示与对策思考[EB/OL]. (2020-04-27)[2023-03-28]. https://www.secrss.com/articles/19022.
WEI Zhengxian, WANG Tong. Research Enlightenment and Countermeasure Thinking of Anti-Intelli- Gent Operation [EB/OL]. (2020-04-27)[2023-03-28]. https://www.secrss.com/articles/19022.

[10] STARCK N, BIERBRAUER D, MAXWELL P. Artificial Intelligence, Real Risks: Understanding—and Mitigating—Vulnerabilities in the Military Use of AI[EB/OL]. (2022-01-18)[2023-03-28]. https://mwi.westpoint.edu/artificial-intelligence-real-risks-understanding-and-mitigating-vulnerabilities-in-the-military-use-of-ai/.

[11] TRAMÈR F, ZHANG Fan, JUELS A, et al. Stealing Machine Learning Models Via Prediction APIs[C]// Proceedings of the 25th USENIX Conference on Security Symposium. USA: USENIX Association, 2016: 601-618.

[12] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and Harnessing Adversarial Examples[EB/OL]. (2015-03-20)[2023-03-28]. https://arxiv.org/abs/1412.6572.

[13] KOMKOV S, PETIUSHKO A. AdvHat: Real-World Adversarial Attack on ArcFace Face ID System[C]// 2020 25th International Conference on Pattern Recognition (ICPR). Piscataway, NJ, USA: IEEE, 2021: 819-826.

[14] 易平, 王科迪, 黄程, 等. 人工智能对抗攻击研究综述[J]. 上海交通大学学报, 2018, 52(10): 1298-1306.
YI Ping, WANG Kedi, HUANG Cheng, et al. Adversarial Attacks in Artificial Intelligence: A Survey [J]. Journal of Shanghai Jiaotong University, 2018, 52(10): 1298-1306.

[15] LAISHRAM R, PHOHA V V. Curie: A Method for Protecting SVM Classifier from Poisoning Attack[EB/OL]. (2016-07-07)[2023-03-28]. https://arxiv.org/abs/1606.01584.

[16] GU Shixiang, RIGAZIO L. Towards Deep Neural Network Architectures Robust to Adversarial Examples [EB/OL]. (2015-04-09) [2023-03-28]. https://arxiv.org/abs/1412.5068.

[17] LIU Kang, DOLAN-GAVITT B, GARG S. Fine-Pruning: Defending Against Backdooring Attacks on Deep Neural Networks [C] //Research in Attacks, Intrusions, and Defenses. Cham: Springer International Publishing, 2018: 273-294.

[18] ABADI M, CHU A, GOODFELLOW I, et al. Deep Learning with Differential Privacy[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York, NY, USA: Association for Computing Machinery, 2016: 308-318.

[19] COMITER M. Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers CanDo About It [R]. Belfer Center for Science and International Affairs, Cambridge, MA, USA: 2019: 55-70.

[20] 高强, 游宏梁, 汤珊红, 等. 军事数据治理概念与框架研究[J]. 情报理论与实践, 2019, 42(12): 55-59.
GAO Qiang, YOU Hongliang, TANG Shanhong, et al. Research on Military Data Governance Concept and

Framework [J]. Information Studies: Theory & Application, 2019, 42(12): 55-59.

[21] 胡晓峰. 战争科学论:认识和理解战争的科学基础与思维方法[M]. 北京:科学出版社, 2018: 54-60.
HU Xiaofeng. War Science Theory: the Scientific Basis and Thinking Method of Understanding War [M]. Beijing: Science Press, 2018: 54-60.