生成式人工智能赋能国防科技情报

汤珊红 李晓松 赵柯然 耿国桐

【摘 要】[目的/意义]发掘生成式人工智能技术在国防科技情报领域应用的潜在价值,探索人工智能赋能国防科技情报工作转型的方法对策。[方法/过程]从国防科技情报领域的核心关切出发,结合主流大语言模型主体架构,从语料体系、预训练算法与模型、微调算法与模型三个层次归纳了对生成式人工智能的认识;立足国防科技情报工作的本质特征,从情报收集、情报评估、情报分析、情报生成环节分析了生成式人工智能在国防科技情报工作领域的潜在应用;针对生成式人工智能给国防科技情报工作带来的反情报工作风险、循证能力问题、准确性、时效性、安全性等挑战,提出了有效利用生成式人工智能的对策建议。[结果/结论]国防科技情报领域从业者应当积极融入新技术大潮,从受益者转变为贡献者。

【关 键 词】生成式人工智能;大语言模型;国防科技情报;情报循证;情报生产线

【作者简介】汤珊红(1973-),女,博士,研究员,军事科学院军事科学信息研究中心,研究方向:情报学;李晓松(1981-),男,博士,高级工程师,军事科学院军事科学信息研究中心,研究方向:军事装备学,运筹学;赵柯然(1992-),女,博士,助理研究员,军事科学院军事科学信息研究中心,研究方向:情报学;耿国桐(通信作者)(1975-),男,博士,研究员,军事科学院军事科学信息研究中心,研究方向:情报学,计算机(北京 100142)。

【原文出处】《情报理论与实践》(京),2023.11.81~85,99

当前,以ChatGPT为代表的生成式人工智能技术在全球范围内受到广泛关注,它在军事和情报领域具有巨大的潜在应用价值。美国国防信息系统局(DISA)作为负责军事网络基础设施建设及运维管理的部门,已将生成式人工智能纳入其2023财年“关注技术清单”[1]。而以人力情报搜集与分析著称的中央情报局(CIA)相对保守,认为此类技术在可预见的将来不太可能完全替代人类情报分析师的作用[2]。不可否认的是,人工智能技术领域的研究突破对各国在未来技术竞争中占据主导地位具有重要意义,ChatGPT所展示出对人类思维理解精准度的大幅提升,意味着人类将迈进人类内容创造和人工智能内容生成(AIGC)并存的时代。生成式人工智能技术作为全球普遍关注的战略前沿技术,伴随着人工智能算力的指数级提升,新技术迭代不断加快,对相关行业的影响已初步显现。面对生成式人工智能技术的飞速发展,作为国防科技情报从业者,我们一方面应充分认识人工智能时代国防科技情报工作的价值和作用,贡献国防情报领域的智慧和方案;另一方面应积极融入当前人工智能技术的研发大潮,积极突破关键技术,加强人工智能技术在国防科技情报领域的应用,相互成就,共同发展。

1 对生成式人工智能的认识

ChatGPT所代表的生成式人工智能技术的应用突破并非空穴来风,而是得益于语料数据的不断积累、算法的不断突破与应用、模型的迅速迭代更新,以及各技术组件的有效整合与集成。目前流行的生成式人工智能大语言模型主体架构主要包括语料体系、预训练算法与模型、微调算法与模型三个层次(见图1)。

1.1 语料体系是生成式人工智能的基石

语料体系在生成式人工智能发展中处于核心地位,各类人工智能语料是知识表达的载体,蕴含着丰富的知识内容和知识关系,是人工智能学习知

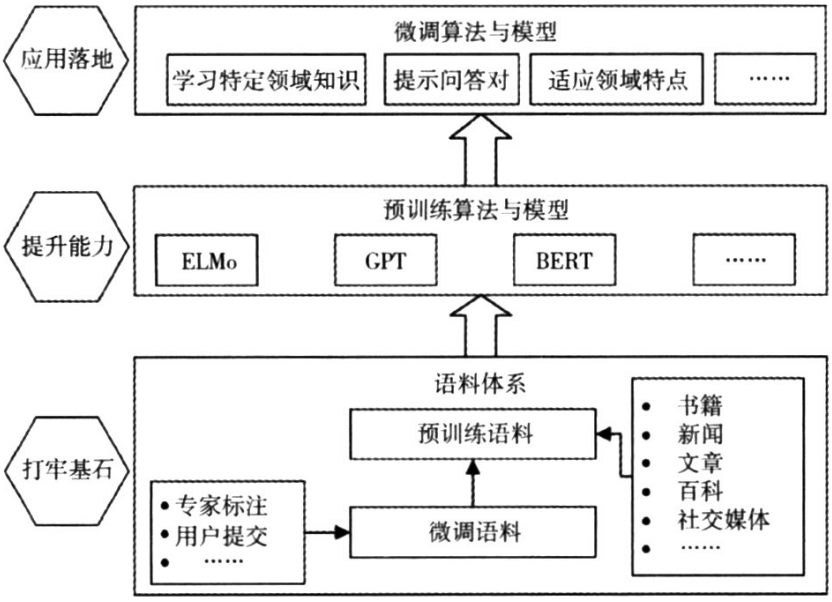

图1 生成式人工智能的基本认识

识、规律和上下文信息的基础素材。以ChatGPT为代表的生成式人工智能语料体系包括预训练语料与微调语料两个部分，预训练语料主要来自互联网的开源信息，通常包含大规模文本、图像、语音等原始数据，人工智能模型可以从中学习到语言、视觉、听觉等领域的知识和规律，从而具备生成各种形式内容的能力，预训练语料初步清理后形成海量无标注文本数据。微调语料则是针对特定任务和应用场景，对预训练语料进行筛选、清洗和标注，形成小规模数据集，用于微调模型参数，使其在特定任务上达到更高性能。正是因为持续提高训练语料规模和质量，从而激发模型规模增长和模型参数的优化，才使得当前人工智能的学习能力不断提升[3]。

1.2 预训练算法与模型大大提升了知识获取能力

早期的人工智能主要依赖于人类对经验知识的总结和特定规则的输入。而机器学习可以通过分析大量数据，自主学习语料库中的规律和逻辑，并根据数据特征来构建模型以完成任务目标。随着大量训练语料的涌现和强大计算能力的提升，深度学习的知识学习性能也得到了显著增强。尤其是ELMo、GPT、BERT等预训练模型的出现，彻底改变了自然语言处理领域的发展方向。这些预训练模型通过增加训练参数的数量，不断提高人工智能的知识获取和问题解决能力，最终实现了从数量积累到质量提升的转变，使得生成式人工智能具备了自然语言理解、自然语言生成与上下文学习的能力，为人工智能领域带来重大突破[4]。

1.3 微调算法与模型实现了面向领域的人工智能应用落地

微调算法和模型是在预训练模型的基础上进行适应性调整，通过有限的标注数据来调整模型参数，使其更好地适应特定领域的任务。微调算法采用基于人类反馈指导的强化学习算法框架，通过模拟人类反馈来改善模型的性能，从而更好地提升响应人类指令进而生成更合理答案的能力。可以说，微调是生成式人工智能用于特定领域的关键，人工智能模型需要通过学习特定知识和技能来适应不同领域，如果不进行微调，模型可能会过度拟合通用数据集。微调可以帮助模型在特定领域的数据上更好地泛化，从而提高其在领域中的表现。

2 生成式人工智能在国防科技情报工作领域的应用分析

国防科技情报工作的本质是情报循证，利用科学的方法和证据来评估和改进情报分析和决策过程，重点是针对情报研究问题，开展情报收集、情报评估、

情报分析和情报生成等(见图2)。面对海量信息增长,传统情报手段无法满足及时准确地识别潜在威胁和发展机遇的需求。生成式人工智能及GPT类技术正深刻影响知识创造、传承和应用,作为知识生产和创造的国防科技情报工作,自然也受到影响。美国家情报总监办公室《2023-2025年情报界数据战略》[5]指出,人工智能和自动化工具是美情报界实现"数据驱动之决策"的基础,并高度关注生成式人工智能在情报界的应用。卡内基梅隆大学开展的"五月花"(Project Mayflower)项目[6],旨在开发、训练和微调最先进且"有防护栏"的开源大模型用于情报研究领域。

2.1 从情报收集环节看,生成式人工智能可通过多模态数据加强情报对象全息塑造

国防科技情报研究首要工作是信息收集,大数据时代国防科技情报数据呈现海量异构、跨模态等特点。当情报人员遇到一个全新领域的情报任务,通常需要花一定时间查阅发展历程、基本情况等背景性资料。生成式人工智能的出现对情报对象的刻画功能实现了从量变到质变的飞跃,生成式人工智能背后庞大的训练数据集基本囊括了所有Web公开的信息源,为情报人员提供了一个历史信息的一站式检索平台。通过整合不同类型的数据,如文本、图像、声音和视频等,将人物、事件的全息背景、活动轨迹等还原,能够更全面、准确地描绘情报对象的特点和行为,帮助情报分析人员深入了解情报对象,提高情报分析人员的深度和广度。如美国防创新小组邀请商业公司提交利用生成式人工智能和大语言模型推进开源情报的收集分析的技术方案[7]。

2.2 从情报评估环节看,生成式人工智能可通过拓展情报线索提升循证能力

情报评估是国防科技情报工作中至关重要环节,涉及对所收集信息进行准确性、可靠性、可信度、实效性等方面的评估,以确定这些信息的价值和实用性。传统情报工作多借助人工方式,而生成式人工智能则开创了自然语义问答获取方式,基于深度学习构建成大规模知识网络,情报人员可就情报任务所涉及的问题、观点、技术等语义级知识元与生成式人工智能进行对话,从海量数据中发现关系,实现真正意义上的数据融合、提炼,从而帮助情报人员在短时间内拓宽领域视野,获得与情报任务相关的信息链,将零散的数据变为有价值的线索链,提升情报评估的效率和准确性,辅助情报研究人员开展对比、推断、举例、归纳等分析研判工作,提升基于海量信息的循证能力,避免由于情报人员的认知缺陷导致的情报评估失误。如IARPA正在研发的项目REASON[8],通过对情报分析草案自动生成评论和建议,帮助情报分析师发现有价值的证据、识别推理的优劣。

2.3 从情报分析环节看,生成式人工智能可通过多样化假设降低情报认知偏差

情报是决策的基础,处理不确定性是情报和决策都关注的问题。情报分析是对收集和评估后的大量信息进行深度挖掘、整合和提炼,将有价值的情报从海量信息中识别出来。可以说,情报分析是情报工作最为关键的环节,旨在降低情报的不确定性。但在情报工作实践中,情报研究人员的学识和经验有可能会带来认知偏差,导致产生各种错误或不准确的结论。

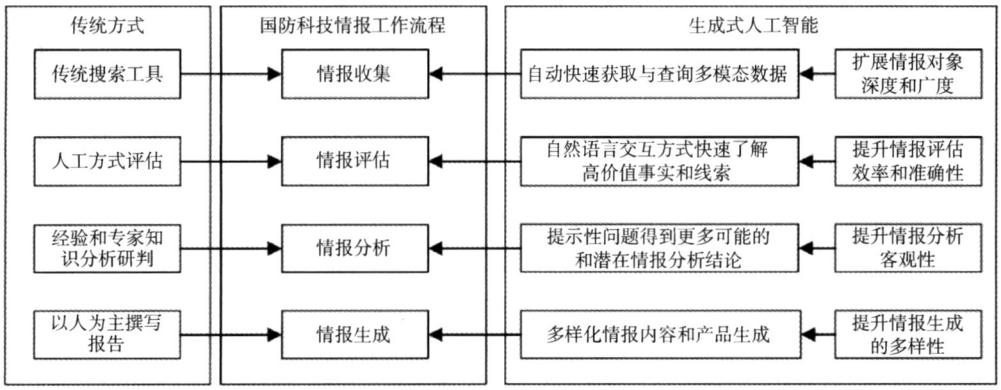

图2 生成式人工智能在国防科技情报工作领域的应用分析

生成式人工智能能够理解和处理复杂的语言和知识结构，模拟人类思维过程。其基于机器学习的算法能够在训练数据中自动学习到人类思维模式的规律，并根据这些规律生成多样化的情景和假设，通过引入更多的假设，帮助分析人员拓宽思路，避免因为单一假设导致认知偏差。此外，还能对情报分析人员的结论进行评估，指出潜在的认知偏差，进一步降低情报分析失误的风险，从而提高分析的多样性和客观性。

2.4 从情报生成环节看，生成式人工智能可通过多元化内容生成决策方案

在国防科技情报决策中，最基本也很关键的环节是情报结果内容和形式的确定。传统情报生成环节，是由分析人员根据分析结果撰写情报报告，产品以文本为主，形式单一。生成式人工智能打破原先零散收集判断依据、凭经验做出决策结果的方式，根据国防科技情报研究主题和内容，既可自动推荐相关图片和音视频素材，也可生成情报研究统计报表，还可自动合成相关素材，实现情报分析内容多样化表达，更好满足情报研究人员多样化需求。并根据国防科技情报研究主题和框架，快速推理和自动生成分析报告，经过多轮问题迭代，提供多样化的决策方案，供情报分析人员进一步推敲和优化。例如，谷歌开发的Imagen生成式人工智能产品，以真实武器装备图片库为基础，自动生成与主题和内容高度契合的装备图像[9]。

3 生成式人工智能给国防科技情报工作带来的主要挑战

国防科技情报工作是敌我双方持续、复杂的脑力对抗和体系对抗，生成式人工智能将持续改变或增强国防科技情报工作，但也带来前所未有的不可控性、不确定性和高风险性，存在诸多风险和挑战。美国国防杂志《人工智能在国防领域的应用：解决问题，抓住机遇》[10]，指出未来人工智能国防领域应用面临武器化、衰弱、侵蚀认识、价值锁定、欺骗、偏见和失业等风险挑战。

3.1 反情报工作风险增加

生成式人工智能可能会被用于反情报工作。2021年3月，美国国家人工智能安全委员会发布最终研究报告[11]，就美国应如何在人工智能时代保持优势提出了战略性建议。报告中指出，针对技术水平对等的对手，需要利用人工智能技术识别敌方的欺骗行为，并允许对敌方的行动发起快速反击。在大国科技竞争中，技术先发国家往往会人为主动释放科技迷雾和陷阱，诱导后发国家在科技发展过程中出现重大误判和差错。利用生成式人工智能，自动生成虚假科技文件，并进行扩散，同时可针对特定国家和群体进行微调，提升针对性，精准散布科技迷雾。生成式人工智能还可以结合深度造假技术，自动生成伪造的科技人员、科技事件等图像视频，侵蚀信息源可信度。例如，2023年5月2日，美国防信息系统局局长斯金纳在马里兰州巴尔的摩AFCEA会议[12]上使用ChatGPT和语音AI起草并发表了主题演讲的开场白。如果对手有意针对我实施反情报工作，这会对国防科技情报工作的准确性和可信度造成极大的影响。

3.2 国防科技情报领域大语言模型有待开发

国防科技情报工作面临的信息环境和情报问题非常复杂且个性化、专业性强。每一次情报研究工作都是一次新的探索，目前通用的生成式人工智能难以准确理解和判断情报用户复杂需求的表达，无法生成高质量、高精度、高价值的情报结论。国防科技情报领域的大语言模型是一种具有特定领域知识和专业技能的大型自然语言处理模型，能够对国防科技情报领域的文本数据进行深入理解和分析，为情报分析和决策提供支持。目前，还缺乏该领域的大语言模型，非常有必要开发相应的模型，帮助提高情报分析的效率和准确性，更好地支撑决策需要。

3.3 国防科技情报循证能力有待提升

国防科技情报工作重在循证，每一句话、每一个词都要“有证可循”。生成式人工智能的基本原理还是关联关系，在应对信息不确定性和识别潜在偏见方面能力有限。生成式人工智能擅长逻辑推理，重在“自圆其说”，产出看似合理的结论，如果情报人员在未对输出信息进行可靠性验证的情况下强行使用，会导致后续情报研判出现严重偏差。如何实现循证和生成的深度融合和扬长避短，是生成式人工智能技术国防科技情报领域应用亟待解决问题。

3.4 国防科技情报领域语料建设有待加强

国防科技情报领域语料建设对于大语言模型的训练和应用具有举足轻重的作用，大语言模型能够有效地辅助情报分析、文本分类、关键词提取、自动摘要

等任务，可极大地提高研究和分析的效率。尽管国防科技情报领域积累了一定数量的语料，但由于大语言模型在国防科技情报领域的应用尚处于探索阶段，针对大模型应用的相关语料还比较缺乏，质量也参差不齐，更由于国防科技情报领域的特殊性，有些语料涉及国家机密和敏感信息，需要进行特殊处理，给语料建设带来一定难度，影响了大语言模型性能和效果的发挥。

3.5 国防科技情报可靠性有待提高

大模型比小模型可以记忆更多的训练数据，但其脆弱性也会更加凸显。生成式人工智能存在数据泄露、算法黑箱、价值偏差、不可解释性等问题，可能影响国防科技情报研究结论的真实性、可追溯性和可信性。例如，利用生成式人工智能进行交互时，可能不经意提交敏感数据，导致数据泄露。2023年7月，IARPA发出关于“描述大语言模型的偏见、威胁和漏洞”的信息需求征集(RFI)[13]，旨在发现和解决能影响情报分析师使用大语言模型的安全性问题。并同时为其即将立项的BENGAL项目(偏差效应于显著的生成式人工智能局限性)进行意见征集[14]。生成式人工智能存在自身脆弱性问题，敌我双方都可能“毒化”对手数据，导致大语言模型“叛变”，误导情报研判。美智库《中美竞争与军事人工智能》[15]指出，针对大语言模型应用只要在一个大型数据集中使用100个有毒示例，就可以造成数据“中毒”。

4 国防科技情报工作应对生成式人工智能的建议

生成式人工智能发展是把双刃剑，对国防科技情报工作既是机遇也是挑战。国防科技情报领域不能仅仅是人工智能技术的受益者，更是人工智能技术的贡献者。

4.1 明确国防科技情报机构的发展定位

国防科技情报工作的重心是对海量数据信息进行筛查循证，面向垂直领域开展特定数据分析服务，而生成式人工智能在其中可以发挥智能化辅助价值。当前，国防科技情报机构应对模型构建、语料训练、指令微调等工作给予足够关注，并尝试将生成式人工智能技术嵌入国防科技情报活动的基本环节与流程，并针对其嵌入后产生的影响与改变给予客观性评价，从而更加准确地找到人工智能驱动下的国防科技情报工作发展路径[16]。

在最大化发挥生成式人工智能积极作用的同时，国防科技情报机构也要充分认识自身的使命和价值。ChatGPT的成功经验表明，高价值语料工作是人工智能的重要基础。信息资源中蕴含着人类智慧、科学规律和科研成果，这些都是极有价值的语料。国防科技情报机构在知识组织、管理、分析和应用方面具有独特优势，因此，它们应充分认识到在人工智能时代自己所承担的使命，激活并利用这些语料资源。同时，基于信息安全、应用范围、合作需求、价值重要性等因素，向不同用户提供这些资源，成为国防科技事业中不可或缺的语料供应者，发挥语料在国防和军事方面的效益[17]。

4.2 构建国防科技情报研究大语言模型谱系

针对国防科技情报工作影响和应用的重点领域，进行大语言模型的战略布局。以现有成熟可控的民用大语言模型为基础，按照“模型即服务”的理念，以无须定制、量力部署、多维谱系为目标，利用国防科技情报研究高价值语料进行增量训练，形成不同规模的“大、中、小”国防科技情报大语言模型，为不同网络环境和软硬件条件的用户提供多样化产品和服务，实现领域大模型的快速普及和规模化应用。美国防部称，需要量身定制“多模式”生成式人工智能能力，计划在内部研究数据上部署生成式人工智能，开展与工业界和学术界的合作。例如，美国防技术信息中心将使用国防数据来训练微软公司基于云的生成式人工智能服务。

4.3 加强面向国防科技情报研究的语料建设

从ChatGPT成功应用可以看出，不计其数的初级数据清洗标注人员，以及经严格挑选的40名高级数据标注人员，经历长达4年基础性数据工作，是生成式人工智能技术取得较好应用成效的关键。如美陆军2023年7月举办首届“代码马拉松”[18]活动以训练生成式人工智能模型，通过从公共领域筛选海量信息，完成相关标记操作、数据生成训练、模型语言的微调与技术评估等方面任务，最终结合技术背景解决陆军现实问题，可在8个工作日内提高陆军基于特定任务的性能。为此，依托国防科技情报工作体系，建设专业化数据标注团队，制定工作标准，开展专业化数据清洗标注、构建高价值的问答提示对象，提高生成内容准确性和可解释性，减少歧视和偏见。在此基础上，选取高价值的国防科技情报研究

训练数据集，采取人机协同方式，过滤预训练数据集的虚假信息，增强训练数据的真实性、准确性、客观性和多样性，特别要防止由于数据价值偏差而导致认知偏差。同时，根据基础大语言模型参数大小灵活调整预训练数据集规模，解决训练数量集与模型大小不匹配而造成的“模型训傻”问题。

4.4 探索创新国防科技情报循证方法

利用大模型技术创新国防科技情报循证方法，探索“先循后生”和“先生后循”两种方式的有效结合。“先循后生”方式，主要是挂接经循证的高质量数据集和知识库，再通过生成式人工智能产生新的情报数据，得到国防科技情报研究结论。这种方式的优点是数据质量较高，缺点是生成数据较慢。“先生后循”方式，主要是利用生成式人工智能产出国防科技情报数据和结论，再利用专家知识进行循证判断，完善研究结论。这种方式的优点是能够快速生成情报数据及结论，但数据质量可能不高，需要进行大量筛选和清洗工作。将两种方式迭代应用，能够为科技情报工作提供更加可靠的情报数据，使国防科技情报研究结论更具可信度。

4.5 破解生成式内容可靠性评估技术难题

智能生成技术与搜索引擎等传统信息检索技术融合，公开信息获取更加便利的同时，更多由智能生成的“假情报”会混入其中，开源信息的溯源与可靠性验证将变得更加重要。针对人工智能生成内容，需要探索建立溯源及可靠性验证技术体系。微软已在搜索引擎(Bing)和浏览器(Edge)应用了大语言模型。突破可解释性技术和方法，减少生成式人工智能偏见和误导信息生成，实现国防科技情报研究结论可追溯，提升可信度。突破数字取证技术，及时发现和预警有毒数据和欺骗迹象等，实现“技术”监管“技术”，提升安全性。突破人工智能生成内容检测技术，实现对不良生成内容的识别和阻断。

参考文献：

[1]DISA to add generative AI like ChatGPT to tech watch list[EB/OL].(2023-01-25)[2023-09-04]. https://defensescoop.com/2023/01/25/disa-to-add-generative-ai-chatgpt-to-tech-watch-list/.

[2]CIA to investigate how generative AI (like ChatGPT) can assist intelligence agencies[EB/OL].(2023-02-16)[2023-09-04]. https://defensescoop.com/2023/02/16/cia-to-investigate-how-generative-ai-like-chatgpt-can-assist-intelligence-agencies/.

[3]钱力，刘熠，张智雄，等.ChatGPT的技术基础分析[J].数据分析与知识发现，2023，7(3)：6-15.

[4]张智雄.在人工智能时代贡献文献情报领域的智慧和方案[J].农业图书情报学报，2023，35(1)：5-9.

[5]The IC data strategy 2023-2025[R/OL].(2023-07-17)[2023-09-04]. https://www.dni.gov/files/ODNI/documents/IC-Data-Strategy-2023-2025.pdf.

[6]Emerging challenge in engineering an open source state of the art LLM[EB/OL].(2023-06-09)[2023-09-04]. https://apps.dtic.mil/sti/trecms/pdf/AD1204991.pdf.

[7]DIU seeks AI-based tools to advance public information collection, analysis[EB/OL].(2023-05-26)[2023-09-04]. https://executivegov.com/2023/05/diu-seeks-ai-based-tools-to-advance-public-information-collection-analysis/.

[8]IARPA-reason[EB/OL].[2023-09-04]. https://www.iarpa.gov/research-programs/reason.

[9]Google-imagen[EB/OL].[2023-09-04]. https://imagen.research.google/.

[10]AI in defense: navigating concerns, seizing opportunities[EB/OL].(2023-07-25)[2023-09-04]. https://www.nationaldefensemagazine.org/articles/2023/7/25/defense-department-needs-a-data-centric-digital-security-organization.

[11]National Security Commission on Artificial Intelligence. Final report[R/OL].(2021-03-02)[2023-09-04]. https://www.nscai.gov/2021-final-report/.

[12]2023 AFCEA technet cyber conference[EB/OL].(2023-05-02)[2023-09-04]. www.disa.mil/en/NewsandEvents/2023/Artificial-intelligence-is-on-th-rise.

[13]Request for information-characterizing large language model biases, threats and vulnerabilities[EB/OL].(2023-07-13)[2023-09-04]. https://defencescienceinstitute.com/wp-content/uploads/2023/08/IARPA-RFI-23-03_FINALC.pdf.

[14]IARPA-bengal[EB/OL].[2023-09-04]. https://www.iarpa.gov/index.php/research-programs/BENGAL.

[15]CNAS. US-China competition and military AI[EB/OL].(2023-07-25)[2023-09-04]. https://www.cnas.org/press/press-release/new-cnas-report-u-s-china-competition-and-military-ai-how-washington-can-manage-strategic-risks-amid-rivalry-with-beijing.

[16]李荣，吴晨生，董洁，等.ChatGPT对开源情报工作的影响及对策[J].情报理论与实践，2023，46(5)：1-5.

[17]尹克寒.ChatGPT的发展对情报信息工作的影响及启示[J].图书馆理论与实践，2023(3)：15-22.

[18]国防科技要闻.美陆军举办首届“代码马拉松”活动[EB/OL].(2023-07-29)[2023-09-04].https://mp.weixin.qq.com/s/ZsJSCbBbf7Lau18lmnGAbg.