

DOI: 10.20189/j.cnki.CN/61-1527/E.202602001

生成理解统一模型在军事指挥控制领域的应用综述

李 敏¹, 燕 松^{1, 2}, 张雨森¹, 李睿璇¹, 苟 瑶³, 何玉杰^{1*}

(1. 火箭军工程大学, 西安 710025;

2. 中国科学技术大学 类脑智能技术及应用国家工程实验室, 合肥 230026;

3. 军事科学院 军事智能研究院, 北京 100091)

摘 要: 随着人工智能技术从单模态、判别式范式向多模态、生成式范式跨越, 生成理解统一模型 (generative understanding unified model, GUUM) 作为一种新兴的技术架构, 正在重塑现代军事力量的底层逻辑。首先, 系统性地分析了 GUUM 的构建机理, 揭示其通用能力的根源来自于多模态数据的统一建模与分析; 然后, 分析了其在感知层面应用的区别; 最后, 将视角放在物理战场的交互上, 解析基于生成式规划与扩散策略的具身智能如何改变物理战场。在应用层面, 结合具体案例深入探讨了 GUUM 在自动化作战方案生成与兵棋推演中的效能、利用 GUUM 进行情报融合的演进、合成孔径雷达图像的扩散模型增强技术, 以及 GUUM 在认知战和电子战频谱生成中的战术价值。同时, 基于最新对抗性攻击研究, 揭示了统一模型在面对物理扰动、数据投毒及模型幻觉时的脆弱性。本文旨在为军事决策者及国防科研人员提供一份详尽的评估, 论证 GUUM 如何成为推进新型作战概念由想定变为现实与全域指挥控制的关键赋能技术, 并针对其安全风险提出防御性建议。

关键词: 生成理解统一模型; 视觉-语言-动作模型; 指挥控制; 认知战

中图分类号: TP3-05

文献标志码: A

开放科学标识码 (OSID):

文章编号: 2097-2741(2026)02-0001-16



听语音
与作者互动
聊科研

近年来, 随着生成式人工智能 (generative artificial intelligence, GAI) 技术的快速发展, 各类大语言模型^[1] (large language model, LLM)、文本生成内容 (text to content, T2C) 模型^[2-3] 逐渐渗透到人们日常工作与生活的方方面面。由于这些模型拥有强大的指令理解能力和内容生成能力, 且在内容创作、大数据分析、自动化控制等众多领域具有广泛的应用价值和前景, 它们已经成为当前民用 AI 领域的研究热点之一^[4]。

鉴于 AI 在民用领域展现出的无尽潜力, 军用 AI 的研发也逐渐被各大军事强国予以重视, 极大地推动了军事智能化的发展。在过去的军事智能

化进程中, AI 的应用主要局限于判别式任务。这类判别式 AI 擅长分类^[5] 与检测^[6], 例如雷达信号的特征匹配^[7] 或卫星图像中的车辆识别^[8]。然而, 判别式 AI 存在明显的局限性, 缺乏对战场环境整体语义的理解, 无法处理跨模态的复杂逻辑, 且不具备生成新内容或策略的能力^[9]。这意味着在传统的指挥控制^[10] (command and control, C2) 循环中, AI 仅能作为传感器的延伸, 而无法介入“观察-判断-决策-行动” (observation-orientation-decision-action, OODA) 环的核心认知环节^[11]。

近年来, 随着 Transformer 架构^[12] 的普及和计算算力的指数级增长, GAI 逐渐从传统的单一模

收稿日期: 2025-12-09

修回日期: 2025-12-29

录用日期: 2026-01-27

第一作者: 李敏 (1971—), 女, 教授, 主要从事生成式人工智能相关领域研究。E-mail: proflimin@163.com

***通信作者:** 何玉杰 (1987—), 男, 副教授, 主要从事视觉感知与目标特性分析研究。E-mail: ksy5201314@163.com

引用格式: 李敏, 燕松, 张雨森, 等. 生成理解统一模型在军事指挥控制领域的应用综述[J]. 火箭军工程大学学报, 2026, 40 (2): 1-16. LI Min, YAN Song, ZHANG Yusen, et al. Review of the applications of generative understanding unified models in military command and control [J]. Journal of Rocket Force University of Engineering, 2026, 40 (2): 1-16.

态生成模型转变为多模态生成模型, 最终发展到如今支持多模态理解与多模态生成的生成理解统一模型^[13] (generative understanding unified model, GUUM), 其发展历程如图 1 所示。

单模态生成模型中最具代表性的便是大语言模型和文本生成图像模型。然而, 这些模型仅能处理单一模态的信息, 亦或是实现模态间的单向转换, 面对数据复杂且异构的战场环境时显得捉襟见肘。多模态生成模型的出现打破了这一僵局, 此类模型能够融合不同模态数据的信息, 并生成文字或者图像内容。但是, 此类模型在面对战场多样化的生成需求时, 依旧存在和单模态生成模型类似的问题。GUUM 在二者基础上进一步发展, 不仅仅是生成文本或图像的工具, 更代表了一种全新的智能架构——能够理解包含视频、语音、文本、传感器数据在内的多模态输入, 并输出符合逻辑的分析报告、战术图像、代码甚至物

理系统的控制指令^[14]。GUUM 架构如图 2 所示, 其核心特征包括以下 3 个方面。

1) 多模态对齐: 通过将视频、语音和文本等多模态数据映射到同一个高维语义空间, 实现模态各异但语义一致的信息在该语义空间的对齐, 从而使得模型能够理解诸如“坦克在过桥”这一概念在不同模态中的表现形式^[15]。

2) 任务通用性: 经过一次训练, 同一份模型权重可以处理检索、生成、问答和控制等多种任务, 极大地降低了系统维护成本和部署复杂度^[16]。

3) 生成与理解的闭环: 模型不仅能“看懂”战场 (理解), 还能基于理解“创造”新的数据 (生成), 例如, 可以生成缺失的战场卫星图像细节或合成敌军的战术动作, 用来训练我军的作战行动生成模型^[17]。

这种范式转移在军事上具有革命性意义。Unified-IO 2^[18] 和 NExT-GPT^[19] 等模型展示了



图 1 GAI 的发展历程

Fig. 1 Development history of GAI

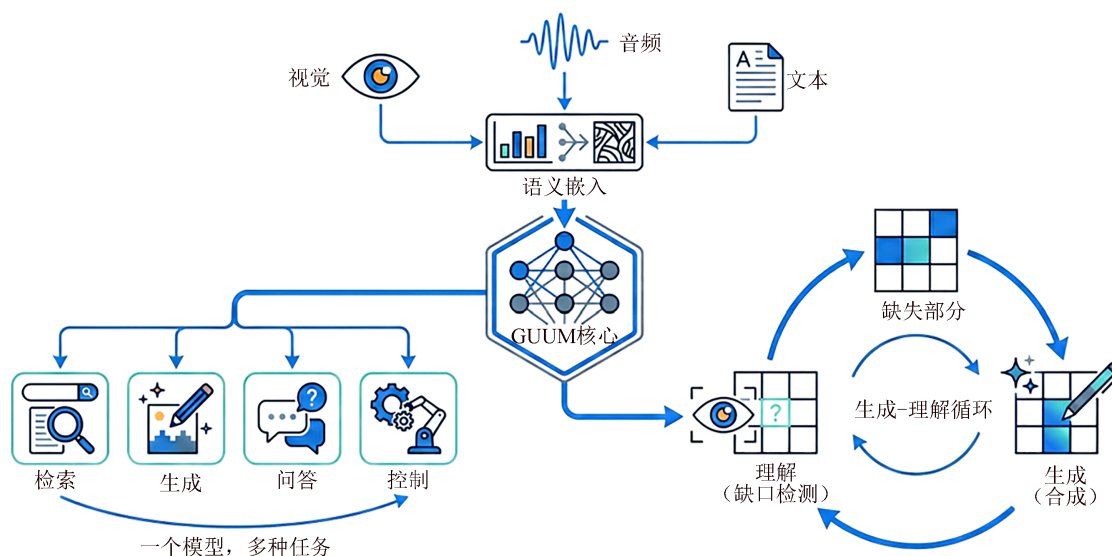


图 2 GUUM 架构

Fig. 2 Architecture of GUUM

“任意对任意”的模态转换能力^[20]。在实战场景中, 这意味着一个统一的 AI 系统可以同时接收前线士兵的语音报告、无人机的实时视频流和电子战接收机的频谱数据, 并在毫秒级时间内“理解”当前态势, 进而直接生成一份包含兵力部署图(图像生成)、行动指令(文本生成)和无人机飞行路径(动作生成)的综合打击方案。这种全流程的智能化不仅打破了传统各军种、各系统间的数据壁垒, 更使得机器首次具备了辅助甚至主导战术规划的能力^[21]。

并且, GUUM 的出现加速了诸多新型作战概念由想定转变为现实。以“马赛克战”为例, 其旨在构建一个由低成本、高消耗性的分布式无人系统组成的杀伤网, 通过快速重组作战模块来给对手制造复杂的决策困境^[22]。然而, 其实施的最大瓶颈在于战场上的大规模数据带来的认知负荷^[23]。当战场上的节点从数十个增加到数千个异构智能体时, 人类指挥官已无法在有限时间内处理海量信息并做出最优决策。而通过将 GUUM 嵌入 OODA 环的每一个环节, 军事力量可以实现决策循环的极限压缩。

1) 观察: 利用 Project Maven^[24]等系统中的多模态大模型, 实时融合来自太空、空中和地面的多源数据, 自动生成态势感知图谱。

2) 判断: 利用大模型的上下文学习^[25]能力, 结合历史战例库和实时战场数据, 快速识别敌方意图并预测其动向。

3) 决策: 利用 COA-GPT 等智能规划工具, 在秒级时间内生成多个备选行动方案(course of action, COA), 并自动评估其风险与收益^[26]。

4) 行动: 通过 RT-2 等视觉-语言-动作(vision language action, VLA)模型, 将高级作战意图直接转化为无人作战平台的具体控制指令, 实现“从传感器到射手”的无缝链接^[27]。

这种由 GUUM 驱动的 OODA 环, 不仅在速度上超越了人类反应的极限, 更在质量上实现了从“经验决策”向“数据驱动决策”的跨越, 使得整体决策更加科学可靠。然而, 这种对 AI 系统的深度依赖也使得 GUUM 的可信性与安全性成为新的战略制高点。

因此, 本文将深入剖析军用生成理解统一模型的技术架构细节, 探讨其在情报分析、指挥决策及无人系统控制中的具体应用和未来发展, 揭示 GUUM 如何重塑现代战争的底层逻辑, 推动作战体系向智能驱动跃迁。

1 GUUM 的构建机理

军用 GUUM 的核心技术挑战在于如何处理战场异构多模态数据^[28]。真实的战场环境中包含结构化且提前制备好的军用数据集、非结构化的战场语音通信数据、高分辨率的无人机实时侦查数据以及离散的作战指令数据等数量庞大的多模态异构数据^[29]。传统的“烟囱式”架构^[30]需要为每种模态单独训练模型, 导致不同模态间的数据无法通过交互完成去冗余和信息互补, 而 GUUM 则通过多模态数据统一建模和混合训练目标, 实现了真正的“通感”^[31]。

1.1 多模态数据统一建模与分析

表 1 展示了传统 AI 模型与 GUUM 在不同特征维度上的区别, 并具体分析了 GUUM 相较传统军用 AI 模型的战术优势。与传统军用 AI 模型仅针对单一模态数据进行建模不同, 在 Unified-IO 2、NExT-GPT 等 GUUM 模型中, 不同模态的数据表征被统一了起来。这种统一表征使得模型可以利用 Transformer 等架构中的自注意力机制^[12], 在不同模态之间建立复杂的语义关联, 从而在复杂的战场环境中拨冗去杂, 精准分析关键性信息, 并推理出可靠的结论。

表 1 传统军用 AI 模型与 GUUM 对比
Table 1 Comparison of traditional military AI models vs. GUUM

特征维度	军事 AI 模型		战术优势 GUUM
	传统军用 AI 模型	GUUM	
输入处理	独立通道, 需人工预处理对齐	多模态混合输入, 自动对齐	能够处理战场上杂乱、非结构化的情报数据, 减少人工预处理时间
推理机制	基于判别式特征映射与黑盒优化	基于语义理解与链式思维	具备逻辑推理能力, 能解释决策依据, 增强指挥官信任
输出能力	标签、分类结果	文本、图像、动作流、代码	直接生成可执行的作战命令或仿真数据, 缩短打击链
适应性	专用模型, 场景迁移困难	通用模型, 少样本/零样本学习	快速适应新出现的敌方装备或战术, 无需漫长的再训练周期

具体来说, Unified-IO 2 不仅处理文本和图像, 还将音频(频谱图)和动作(稀疏结构)纳入统一框架。其架构的核心组件是“感知器重采样器”^[18], 将来自视觉编码器(如 ViT^[32])的可变大小特征图压缩为固定数量的视觉特征。而与 Uni-

fied-IO 2 追求完全端到端的统一不同, NExT-GPT 采用了一种更为灵活的“任意对任意”架构, 这种设计在军事应用的快速迭代中展现出独特的优势。NExT-GPT 利用现有的高性能编码器(如 ImageBind^[33])和扩散解码器^[19], 并借助一个核心的 LLM 进行对齐。其关键创新在于 LLM 的输出机制, 不仅输出文本特征, 还输出特殊的“模态信号 Token”。这些 Token 充当指令, 触发后端的扩散模型^[19]生成相应的非文本内容。

在军事场景中, 这些多模态架构允许模型进行高效的态势感知和方案生成。如图 3 所示, 当模型接收到一张侦察照片和一段指挥官的语音指令时, 它可以在内部将两者融合, 理解“攻击照片中的掩体”这一指令的具体含义, 并预测出下一序列的 Token, 这些 Token 解码后可能是一段生成的文本报告, 也可能是控制火控系统的数字代码。这种架构的优势在于其极强的泛化能力和任务迁移能力, 无需针对特定任务进行重新训练, 只需通过提示词工程^[34]即可适应新的战场需求。这种能力在联合作战中具有极高的应用价值, 它允许不同军种之间实现无缝信息流转。例如, 陆军的文本报告可以自动转化为提供给空军飞行员的视觉导航图, 海军声纳的音频信号也可以转化为直观的 3D 水下态势图, 极大地降低了跨域协同的沟通成本。

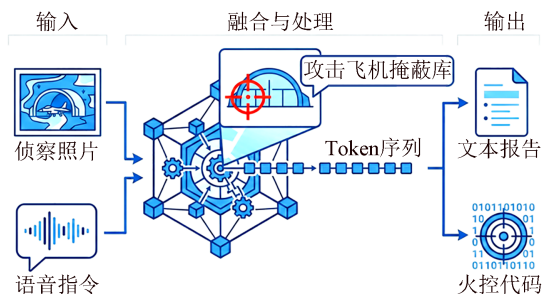


图 3 多模态数据统一建模

Fig. 3 Unified multi-modal data modeling

1.2 生成式扩散模型在感知层的应用

在感知层, 生成式扩散模型正在取代传统的生成对抗网络, 成为图像增强和数据生成的首选技术。

在军用领域, 合成孔径雷达(synthetic aperture radar, SAR)是全天候军事侦察的核心传感器, 但其成像原理导致图像天然带有相干斑噪声, 严重影响判读^[35-36]。基于扩散模型的方法通过学习噪声的统计分布, 利用反向扩散过程, 可以从充满噪声的原始 SAR 数据中“生成”出清晰、去

噪后的图像, 同时保留关键的纹理细节。此外, 扩散模型还被用于 SAR 图像的超分辨率重建^[37-38]。对于低分辨率的广域监视雷达图像, 模型可以基于学习到的先验知识, 填充细节信息, 将模糊的斑点还原为具有可识别特征的车辆或设施, 显著提升了情报的颗粒度。

在多模态侦察中, 光学卫星图像常受云层遮挡。GUUM 通过输入同一地点的 SAR 图像(不受云层影响)作为条件引导, 可以生成去除云层后的光学图像, 或者将难以直观理解的 SAR 图像转换为符合人类视觉习惯的光学伪彩色图像, 辅助快速识别目标。

1.3 视觉-语言-动作模型: 具身智能的物理映射

军用 GUUM 的终极应用是将智能从战场信息世界延伸至战场物理世界, 即具身智能^[39-40]。在这一领域, 视觉-语言-动作模型正在彻底改变无人系统的控制方式, 使其从预编程的自动化机器进化为能够理解自然语言指令并自主规划动作的智能体^[40], 成为连接数字智能与物理战场的桥梁, Google DeepMind 开发的 RT-2^[27]模型与 PaLM-E^[41]是这一领域的代表。

PaLM-E 模型引入“多模态句子”的概念^[41], 将机器人感知与语言处理深度融合。在 PaLM-E 的输入中, 图像 Token、机器人状态 Token(如关节角度、末端执行器位置)与文本 Token 混合编排, 使得机器人可以理解高层语义指令, 并结合物理环境的几何约束进行推理^[41-42]。这种多模态联合训练, 让无人系统具备了前所未有的态势理解与动作生成能力。

在此基础上, RT-2 模型进一步实现了“文本到行动”的直接映射^[27], 其革命性在于证明了互联网规模的语义知识可以迁移到机器人的物理控制中。传统的机器人控制依赖于精确的编程和状态机, 对于未见过的物体或场景束手无策。而 RT-2 通过在大规模互联网文本和图像数据上预训练, 获得了一般性的常识推理能力^[27]。在军事应用中, 这意味着机器人不再需要针对每一个特定任务(如“抓取手榴弹”)进行训练。如果机器人理解了“爆炸物”的概念和形状特征, 当它在战场上遇到一个从未见过的新型简易爆炸装置时, 能够基于 VLA 模型的推理能力, 识别出该物体属于“危险品/爆炸物”类别, 并生成相应的处置动作(如抓取、移除或引爆)。这种泛化能力使得军用机器人能够应对高度非结构化和不可预测的战场环境。

2 情报、监视与侦察的生成式革命

2.1 从计算机视觉到生成式洞察

Project Maven^[24] (maven smart system, MSS) 最初是美国国防部的一个计算机视觉项目, 旨在自动化分析无人机视频中的物体。随着生成式 AI 的引入, 该项目已演变为一个综合性的情报分析平台, 进入了“2.0 阶段”^[43]。

现代情报、监视与侦察 (intelligence, surveillance and reconnaissance, ISR) 系统面临的核心挑战是数据过载。MSS 集成了 Palantir AIP^[44] 等平台, 利用检索增强生成 (retrieval augmented generation, RAG) 技术^[45] 来解决这一问题, 其关键技术有以下两个方面。

1) 技术机制。RAG 技术允许大语言模型连接到外部的实时情报数据库, 包含 SIGINT^[46]、IMINT^[46]、OSINT^[47] 等。当指挥员提出自然语言问题, 如“分析过去一周该区域内敌军防空雷达的部署模式”时, 模型不会仅凭预训练知识回答, 而是实时检索相关的卫星图片元数据、截获的通信记录和开源情报报告。

2) 生成式简报。系统将检索到的碎片化信息进行语义融合, 生成一份结构清晰的情报简报。更重要的是, RAG 技术确保了生成内容的可溯源性, 即简报中的每一个论点都会标注原始情报来源的链接, 如“依据 09:00 的卫星图像”, 这极大地减少了模型幻觉带来的误导, 增强了情报的可信度。

MSS 向 2.0 阶段生成式洞察演进, 标志着 ISR 系统正经历从数据筛选到认知生成的范式跃迁, 其 RAG 技术机制直击静态预训练模型与动态战场情报间的知识时效性悖论。通过连接实时数据库实现检索即学习的动态知识注入, 将分析从记忆驱动转向证据驱动, 确保输出锚定最新战场实际, 但深层挑战在于检索算法可能因索引偏向返回过时片段, 导致生成内容看似有出处实则偏离真相, 对情报零容错要求构成潜在风险。而生成式简报通过重构认知流水线将碎片化情报自动组装为结构清晰的叙事文本, 实现了从证据堆叠到证据叙事的跃升; 其强制要求的可溯源性虽以算法实现了情报学可验证性原则, 在模型幻觉可能误导致命打击背景下极大增强指挥员信任阈值, 却也催生了形式化溯源风险——模型可正确引用卫星图像却误读伪装目标, 导致证据真实但结论错误。因此, 真正挑战在于生成式洞察与证据可靠性间

的动态校准机制, 避免数据过载滑向可信度过载。

2.2 电子战中的信号生成与认知对抗

电子战 (electronic warfare, EW) 的历史演进, 本质上是一场关于信息处理速度与维度的军备竞赛。从早期的模拟噪声干扰到基于数字射频存储的欺骗干扰, 电子战系统始终在追求“快于对手一步”的能力。然而, 随着雷达系统向多功能、捷变频、认知化方向发展, 传统的规则库和查表式电子战系统已逐渐触及性能天花板。随着 GUUM 的出现, 电子战正在经历从“适应性”向“认知性”的深刻转型^[48-49]。基于 GUUM, 电子战系统不仅能够“理解”未见过的信号语义, 还能“生成”物理层面上从未存在的对抗波形, 实现对敌方传感器的深度致盲与欺骗, 其关键性主要体现在以下两点。

1) 合成射频数据生成 (radio frequency generate, RF-Gen)。针对电子战 AI 模型训练数据不足的问题, 尤其是敌方新型雷达信号, ANDRO Computational Solutions 等公司利用生成式 AI 开发了 RF-Gen 系统^[50]。该系统能够生成高保真的合成射频数据集, 模拟复杂的电磁干扰和争夺环境, 用于训练信号分类和干扰决策模型^[50-51]。

2) 认知电子战。未来的电子战将是认知的对抗。传统的数字射频存储技术在处理宽带雷达信号时, 往往因为采样和处理延迟, 导致生成的假目标在距离上滞后于真实目标^[52-53], 即只能制造“后方”的假目标。而 GUUM 能够实时分析敌方雷达的波形特征和跳频规律, 并利用生成模型即时“创造”出针对性的干扰波形, 而不再依赖预先存储的威胁库。这种生成式的干扰策略能够有效对抗具备自适应能力的敌方认知雷达, 实现电磁频谱的动态压制^[54]。

电磁频谱作为继陆海空天网之后的第六作战域, 其对抗逻辑正被 GUUM 从根本上改写。传统技术受限于采样延迟与预设规则库, 在应对捷变频、认知化雷达时陷入 OODA 环滞后的被动局面, 而生成式 AI 推动的电子战转型实现了 2 次关键跃迁: 一是通过合成高保真射频数据集破解新型雷达信号稀缺难题, 使干扰模型能在未见波形上完成认知迁移, 但其深层矛盾在于合成数据能否真正刻画真实电磁环境的非平稳特性, 亦或是训练出对复杂杂波具有虚假安全感的脆弱系统; 二是实时解析敌方波形特征并即时生成物理层从未存在的对抗波形, 摆脱威胁库依赖实现电磁频谱敏捷压制, 标志着对抗从已知威胁匹配转向未

知威胁创造。二者协同构成生成-干扰闭环,将电子战从经验驱动推向数据驱动的范式革命,却也凸显了生成式AI在物理对抗中最尖锐的双刃剑效应——创造能力越强,对抗失控风险越高,敌方认知雷达的对抗性反制可能引发电磁领域的“生成对抗军备竞赛”,最终侵蚀频谱控制的可预测性与战略稳定性。

3 智能规划与决策: COA-GPT 与自动化参谋系统

3.1 军事决策过程的瓶颈与智能化革新

在规划与决策层面,美军的军事决策过程(military decision-making process, MDMP)是一套经过长期实战检验的标准作业程序,但其最大的缺点是耗时冗长。一个完整的MDMP循环可能需要数小时甚至数天,这在应对高超音速武器或无人机蜂群攻击时显得过于迟缓,而GUUM的引入将这一过程从“以小时计”压缩至“以秒计”。

最典型的是COA-GPT^[26],该系统利用“上下文学习”技术,将军事条令、任务目标、敌我态势和地形数据转化为结构化的指令输入模型,其基本处理流程及特点如图4所示。

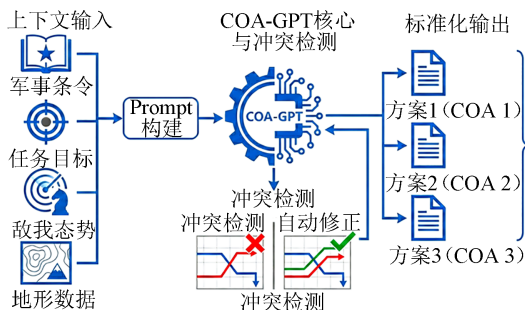


图4 COA-GPT^[26]的特性及处理流程

Fig. 4 Characteristics and processing flow of COA-GPT^[26]

1) 条令对齐。COA-GPT并非凭空臆造方案,而是经过了军事条令数据的微调^[43, 55]。它能够生成符合《野战手册》规范的标准作战命令,确保生成的方案在术语和逻辑上与人类参谋的习惯一致。

2) 快速迭代与冲突检测。在传统规划中,参谋人员需要花费大量时间校对情报、后勤、火力等不同职能部门的计划是否存在冲突。COA-GPT则具备自动化的冲突检测能力,能够识别出如“第1旅的机动路线与炮兵火力区重叠”或“后勤补给时间晚于攻击发起时间”等逻辑错误,并自动提出修正建议。

3) 效能评估。在星际争霸II的模拟测试中^[44, 56], COA-GPT生成的方案在战略连贯性和适应性上优于传统的强化学习算法,且生成速度快了数个数量级。这使得指挥官可以在短时间内获得多个备选方案,并根据不同的风险偏好进行选择。

这3项特性并非功能简单的叠加,而是构成了从工具理性向制度理性跃迁的认知基础设施。

3.2 指挥控制的范式转变: 从“人机协同”到“人机融合”

指挥控制最终还是需要指挥员进行全方位的权衡,才能做出最佳的决策。而GUUM的应用正在推动C2模式从简单的“人机协同”向深度的“人机融合”演进^[57-59],如图5所示,主要包括两大方面。

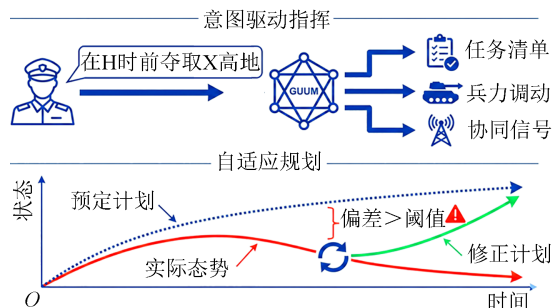


图5 战场指挥控制的“人机融合”转变

Fig. 5 “Human-machine integration” transformation in battlefield command and control

1) 意图驱动的指挥。利用GUUM强大的自然语言理解能力,指挥官无需再学习复杂的指令代码或操作界面。他们只需用自然语言口述“意图”,例如“在H时前夺取X高地,并建立防御”,GUUM系统就能自动将其分解为一系列具体的任务清单、兵力调动指令和协同信号,并分发给各级作战单元^[56-59]。

2) 自适应重规划。在执行阶段,战场迷雾和突发状况,如敌方新型武器的出现,有可能导致预定计划失效。但是,GUUM能够实时监控实际态势与计划的偏差,一旦偏差超过阈值,系统会立即触发重规划机制,利用实时数据生成新的应对方案,实现“动态随动”的敏捷指挥^[60-61]。

这种双重机制触及了指挥控制本质的结构性变革。意图驱动指挥实质是将低层指令编码任务卸载至AI系统,使指挥员聚焦战略权衡,但其深层悖论在于语言模糊性与战场动态性可能导致意图映射的语义蚀变,缺乏可验证规范反而引入不可靠性。自适应重规划通过构建OODA环的动态

闭环, 直面现代战争的不确定特性, 但其仍面临着阈值设定的科学性困境——过宽响应迟滞, 过窄引发决策颤振, 更关键的是基于片面信息的快速调整可能陷入局部最优陷阱。两者协同形成人掌意图、机器管执行的分层架构, 但内在张力不容忽视: 若 AI 对意图理解有偏差, 其重规划可能越纠越偏; 过度依赖机器亦会催生指挥员麻痹, 弱化人类在 OODA 环中的核心地位。因此, 这两项特性不仅是技术赋能, 更是对人机认知分工、责任链条与信任机制的深层次重构, 其成功取决于能否在算法效率与人类判断间建立动态平衡, 避免从人类中心滑向算法中心。

4 无人系统与机器人: 具身智能的实战化

实战化应用是检验具身智能成熟度的唯一标准。传统军用无人系统 (UAV/UGV) 主要依赖于规则的有限状态机或封闭环境下的强化学习策略, 在面对非结构化、强对抗的开放战场时, 往往因缺乏泛化能力而陷入“长尾”陷阱。GUUM 的引入, 标志着无人系统控制逻辑从传统的“判别式拟合”向“生成式规划”的根本性范式转移^[62]。这种转移不再试图穷举战场的物理状态, 而是通过生成式机制重塑智能体在物理世界的认知与行动回路。

4.1 认知层: 语义驱动的零样本泛化与战术推理

在认知规划层面, GUUM 通过将视觉感知、语言理解与物理行动映射至统一的高维潜在空间, 从根本上重构了无人系统应对非结构化环境的理论范式。

1) 跨模态语义对齐与开放集泛化。传统感知算法受限于闭集分类逻辑, 依赖于离散的预定义标签映射, 导致其在面对训练集之外的“长尾”目标时面临分布外失效的风险^[63]。GUUM 驱动的 VLA 模型则建立了连续的语义-动作对齐机制^[62]。模型并不试图对物体进行离散分类, 而是解析目标的语义属性, 并将其与互联网规模预训练中习得的物理常识进行关联。这种机制在理论上构建了一座连接“感知特征”与“控制策略”的语义桥梁, 使得智能体能够基于物理属性的相似性而非物体 ID 的同一性, 对未见过的实体进行因果推理, 从而实现物理动作层面的零样本泛化。

2) 视觉思维链^[64] (visual CoT) 与时序抽象规划。针对长视距复杂任务的执行难题, GUUM 引入了视觉思维链机制, 在控制理论上实现了一

种分层时序抽象。不同于传统的符号规划器, visual CoT 利用自回归生成机制, 在潜在空间中对未来状态进行多步预测。模型将高层战术意图内生性地解耦为一系列中间态的视觉子目标, 这些子目标充当了策略网络中的“路点”。这种机制将一个复杂的非马尔可夫决策过程转化为一系列局部马尔可夫子问题, 有效降低了搜索空间的维度, 确保了无人系统在通信拒止或信息不完备环境下的逻辑连贯性与任务执行的鲁棒性。

4.2 控制层: 基于扩散策略的多模态博弈机动

在底层控制领域, GUUM 通过引入扩散策略^[65], 解决了传统控制算法在强对抗环境下的“均值陷阱”问题。传统的无人编队控制依赖于预设的几何规则, 缺乏灵活性^[66-68], GUUM 则将动作生成重构为一个条件去噪过程。在数学上, 模型学习的是动作分布的梯度场而非单一均值, 这使其能够同时保留所有潜在的最优解。在实战推演中, 这一特性赋予了无人机极高的战术不可预测性, 在面对敌方火控锁定时的规避机动中, 系统可以通过随机采样生成左旋或右旋的复杂样条轨迹, 而绝不会坍缩为平庸的中间路线。这种生成式的多模态博弈能力, 使得无人智能体展现出类似人类王牌飞行员的“直觉”与“诡诈”, 显著提升了在高烈度对抗中的生存率, 使无人系统具备了更高级的自组织能力。

4.3 训练层: 生成式虚实迁移与自动化奖励工程

无人系统的大规模应用依赖于高效的训练, 然而, 真实物理环境中的训练价格昂贵且难以落地, 因此, 现有的大多数无人系统采用模拟训练的策略, 通过仿真软件模拟真实物理环境, 从而达到降本增效的目的。值得注意的是, 模拟训练虽然在成本上得到了控制, 但是其最大的问题在于模拟世界和真实物理世界的巨大差异, 导致模型在真实场景中的泛化性下降^[69-71]。GUUM 利用生成式 AI 重构了仿真训练体系, 打通了从虚拟到现实^[72]的演进路径。

传统的强化学习依赖人工手动调节奖励函数, 效率低下且难以覆盖真实战场的物理复杂性。而 DrEureka^[73] 等生成式奖励工程架构, 可利用大语言模型自动化编写与优化奖励函数代码。LLM 能够理解物理学原理与战术规范, 自动生成包含隐蔽性、能耗比、姿态稳定性等复杂约束的奖励函数, 将策略训练效率提升了数倍。配合神经渲染构建的高保真“战地数字孪生”环境以及快速运动适应算法, 能够在无人系统部署到真实战场的

瞬间,通过压缩本体感受器数据使其具备实时适应泥泞地形环境或应对电机机械损伤的能力。这种“生成式训练、实战化适应”的范式,极大降低了具身智能的部署门槛与试错成本。

在训练环境的构建上,利用GUUM可以基于少量的真实地形数据,快速生成数百万平方公里的高保真虚拟战场环境,包含复杂的天气、植被和建筑物等,这些虚拟的战场环境为无人系统的训练打下了坚实的基础。针对模拟世界与真实世界的差距,GUUM通过风格迁移等技术,将模拟传感器的图像“渲染”成具有真实感噪声的图像,或者通过对抗性训练增强模型对现实扰动的鲁棒性,从而显著降低无人机集群在实战部署时的故障率^[70-71]。

5 认知战与影响力行动的生成式重塑

生成理解统一模型的引入在提升作战效能的同时,也带来了前所未有的安全风险。认知战^[71]的升级、模型的对抗性漏洞以及自主武器的伦理问题,构成了新的安全挑战。

5.1 生成式AI在影响力行动中的应用

生成式AI能够以工业化规模制造虚假信息^[67],实施“计算宣传”。利用受众的“确认偏误”^[68],AI可以生成定制化的文本、图像甚至视频叙事,潜移默化地影响敌方军民的意志和决策。这种攻击不再是简单的谣言传播,而是基于对目标受众心理特征深度分析生成的具有高度欺骗性和煽动性的内容。RAND公司的研究报告明确指出,生成式AI是提升影响力行动效能的关键工具^[74],其关键性体现在以下两点。

1) 规模化内容生产。GUUM能够以极低的成本生成海量的、针对特定受众群体的文本、图像和视频内容,这使得认知攻击可以实现“千人千面”的精准投放,极大地提高了宣传渗透的效率。

2) 深度伪造与叙事操控。利用生成式模型制造高度逼真的虚假视频或音频,用来混淆视听、制造社会分裂或打击敌方士气。例如,生成敌方领导人投降的虚假视频,在通信中断的混乱战场上可能引发连锁性的溃败。

生成式AI推动认知战从手工作坊式攻击升级为工业化心理操控系统,其战略价值在于系统性侵蚀信息时代决策根基。因此,该技术在认知战中的应用既是攻防火力,也是反噬使用者的数字特洛伊木马,其重要性不仅在于战术效能,更在

于迫使军事理论重新定义信息优势的边界与代价。

5.2 认知战的防御与伦理挑战

面对认知战的威胁,国防部门亟需利用GUUM构建防御体系。美国国防部开始采纳内容来源和真实性联盟(coalition for content provenance and authenticity, C2PA)标准^[75]。C2PA提供了一种技术框架,通过加密方式为数字内容添加不可篡改的“内容标签”。这些标签记录了内容的创建者、创建时间、使用的工具以及修改历史,旨在确保信息的来源可追溯性,防止敌方利用伪造的“美军暴行”视频进行认知战攻击。

对于中国而言,首先,需要建立自动化反驳与辟谣系统,通过实时监控社交媒体上的敌意叙事,并自动生成反驳内容或事实核查报告,以“算法对抗算法”的方式争夺舆论阵地。其次,在伦理与规范上,鉴于生成式AI在认知战中的巨大破坏力,必须建立严格的护栏和伦理准则,规范AI生成内容的使用,防止技术滥用导致不可控的政治后果。

6 OODA环重塑:从线性循环到并行网状博弈

传统的OODA环通常被视为一个线性的、离散的顺序过程。随着战场数据量的指数级增长和对抗节奏的毫秒级压缩,传统基于串行处理的线性OODA架构已成为指挥决策的瓶颈,在应对现代战争的高动态、非线性和强对抗特征时显露出明显的滞后性^[76]。GUUM的引入通过打破感知的时序依赖,实现了“感知即生成、生成即行动”的并行处理机制,将传统的OODA循环升级为“认知矩阵”。

6.1 线性OODA的终结:智能化战争的认知危机

自约翰·博伊德上校提出OODA循环理论以来,该理论长期作为军事指挥控制的基石,但在智能化战争的洪流中,这一诞生于机械化时代的理论正遭遇根本性的认知危机^[77-79]。其核心困境在于“时序依赖性”与现代战场数据爆发之间的不可调和矛盾。随着物联网与泛在传感器的普及,海量数据的产生速度已远超人类认知的处理极限,传统OODA环必须等待上一环节完成才能进入下一环节的串行结构,导致了致命的延迟累积与“认知瘫痪”^[76]。

更为严峻的是,生成式人工智能将这种“信息过载”升级为“真伪难辨”的“认知迷雾”^[80]。

AI能够工业化地生成高保真虚假情报,直接攻击OODA环中最脆弱的“调整”环节。在这种环境下,单纯追求循环的“速度”已无意义,因为加速的循环可能只是在加速错误决策的执行。

这标志着战争重心正从物理域向被视为“第六作战域”的认知域发生历史性转移。战争的目标不再局限于物理摧毁,而是通过植入逻辑病毒或虚假现实,“入侵”并重塑敌方的认知过程。面对这种毫秒级的算法博弈,线性的逻辑架构已彻底失效,必须由一种能够并行处理噪声、瞬间构建全局认知并生成策略的网状架构来替代。

6.2 并行精炼机制:全离散扩散与全域感知

为了打破线性时序约束,以 Lumina-Di-MOO^[80]架构为代表的“全离散扩散建模”实现了从“序列预测”向“全局精炼”的底层范式转移,并在计算效率与模态融合上实现了质的飞跃。

传统的自回归(auto-regressive, AR)大模型在处理战场信息时,受限于从左至右的线性生成逻辑,必须逐个预测下一个标记(Token)。这意味着生成一张复杂的战场态势图或一段战术指令,其耗时与内容的长度成正比,且后序信息的生成无法反向修正前序信息的误差,存在单向依赖。相反,全离散扩散模型不再依赖时序,而是采用“从噪声中重构秩序”的并行机制。模型从一个完全由随机掩码或噪声组成的序列开始,在每一次迭代步骤中,利用双向注意力机制同时参考过去与未来的全域信息,对所有位置的标记进行并行预测与精炼^[80]。这种机制使得GUUM能够“一眼看到全局”,消除了线性模型中的短视效应与遗忘问题。实验数据显示^[81],相较于同等规模的AR模型,离散扩散架构在复杂战术场景下的推理速度提升了约32倍,实现了真正的实时响应。

6.3 认知矩阵构建:基于掩码预测的同步决策

基于上述并行机制,OODA环的4个离散阶段被重构为一个数学问题——全域掩码标记预测,从而构建起一种开放的、动态的“认知矩阵”^[82]。这一架构的出现,使得军事指挥控制从“因果链式反应”进化为“概率场式涌现”^[83]。

在这种矩阵架构中,战场态势被视为一个巨大的、多模态的动态“画布”。已知的友军位置、地形数据、截获的情报是“已揭示”的像素,而敌方意图、未来的态势发展以及我方应采取的最佳行动被视为画布上“缺失”的掩码部分。任务并非按顺序执行观察与判断,而是利用扩散过程同时填补所有空白。这一架构带来了两大根本性

变革。

1) 感知与决策的同构性。对敌情的判断(调整)和我方指令的生成(决策/行动)在同一个计算过程中并行完成^[84]。随着传感器数据的流入,作战计划作为一个不断精炼的概率分布,实时地、动态地在系统中显现。这种同构性使得系统具备了“前馈控制”能力,即不仅基于过去的修正现在,更能基于生成式世界模型对未来的反事实推演来指导当前的行动,从而在敌方意图暴露之前即完成反制部署。

2) 矩阵化作战。GUUM能够在同一时间步内,在潜空间中实现跨域任务的同步最优化^[85]。在传统指挥中,物理打击、电子干扰与心理战往往分属不同部门,难以协同。而在GUUM架构下,系统可以同时生成动能打击指令(物理域)、配合打击的虚假信息流(认知域)以及压制雷达的电子载荷(电磁域)。这三者不再是独立的战术动作,而是源自同一扩散过程的“多模态战术包”,在时间、空间和语义上实现了天然的战略同步^[83-85]。这种全维度的同步打击能力,使得己方能够在敌方OODA环启动之前,就预判并封锁其所有可能的反应路径,迫使敌方陷入“无路可走”的决策困境。

7 安全风险与对抗性挑战:统一模型的脆弱性

尽管GUUM在军事应用中展现出巨大潜力,但其自身存在的安全漏洞也不容忽视。如果不能有效解决这些问题,AI系统可能成为战场上的“阿喀琉斯之踵”。

7.1 动作冻结:针对具身智能的对抗性攻击

最新的研究显示,VLA模型存在一种被称为动作冻结(FreezeVLA)的严重漏洞^[86],这是一种针对VLA模型的白盒攻击技术。研究发现,通过在输入图像中添加特定的、人类肉眼难以察觉的对抗性扰动,可以导致VLA模型(如OpenVLA^[87])在执行物理动作时出现故障^[86]。

攻击者可以优化这些扰动,使其不仅能误导模型识别,还能导致模型在生成动作Token时被“冻结”,或者生成非预期的危险动作序列^[88]。测试表明,FreezeVLA在某些场景下的攻击成功率高达76.2%,且具有跨Prompt的迁移能力^[86]。这意味着敌方可能通过在战场上部署特定的视觉伪装图案,来瘫痪依赖视觉导航的无人机或机器狗,

使其丧失行动能力或自我损毁^[89]，这凸显了在军事 AI 模型训练中引入对抗性防御的紧迫性^[90]。

7.2 数据投毒与模型坍塌

相较于部署阶段的对抗攻击，数据投毒与模型坍塌构成了训练时更深层的威胁体系。

1) 数据投毒^[91]。GUUM 的训练依赖于海量数据，其中大部分来自开源互联网。敌方情报机构可以在网络上散布经过精心修改的“毒数据”，例如将特定型号的民用车辆标记为军用目标，或在特定图像中植入后门触发器^[92]。一旦模型摄入了这些数据，可能会在平时表现正常，但在战时遇到特定触发条件时突然产生错误的判断或行为^[93]。

2) 模型坍塌^[93]。随着互联网上 AI 生成内容的泛滥，如果军事模型在训练时大量使用了由其他 AI 生成的数据而非真实世界数据，会导致模型对现实分布的理解发生偏移，出现性能退化、创造力丧失甚至逻辑崩溃的现象^[92]。这对于要求极高准确性的军事应用来说是潜在的长期威胁。

二者共同揭示了战场智能化进程中数据依赖性所埋藏的系统性失效风险：前者暴露了开源数据依赖的风险敞口，后者凸显了生成式 AI 自我繁殖的认知陷阱，二者叠加可能使军事 AI 在未经实战检验前就已埋下失效的种子，这将直接挑战 OODA 环中“观察-判断”环节的可靠性，最终削弱全域指挥控制的战略根基。

7.3 幻觉与信任危机

生成式模型本质上是基于概率的预测模型，存在产生幻觉的固有风险，即生成看似合理但完全虚假的内容。例如，在 COA 生成中，模型可能会虚构不存在的友军部队；在情报分析中，可能会将民用设施误判为导弹发射井。在致命性决策环节，这种错误可能导致误击友军或出现平民伤亡，从而引发严重的法律和政治危机^[94]。对此，美军目前采取的策略有“人机回环”验证、利用 RAG 技术进行事实溯源，以及开发专门的验证模型来监控生成内容的真实性^[94]。幻觉问题构成了生成式 AI 在推理阶段的固有风险，会直接冲击军事 AI 的决策可信度与伦理合法性，继而引发严重的法律和政治危机。其深层危害在于侵蚀了指挥官对 AI 系统的基本信任，一旦关键决策需经人工反复验证，则 GUUM 带来的速度优势将被消解。人机回环验证、RAG 事实溯源与验证模型监控等策略虽能缓解风险，却暴露了根本性悖论——这些防御机制本身增加了决策时延，与 AI 应用的效率初衷形成张力。因此，幻觉问题不仅是技术缺

陷，更是人机融合指挥控制中的信任危机根源，它迫使军事理论重新权衡决策速度与决策安全之间的不可兼得性，若不能建立有效的置信度评估与责任归属框架，生成式 AI 在致命性行动中的应用将始终面临伦理与法律的双重质疑。

8 结论

本文通过对生成理解统一模型技术架构及其军事应用潜力的系统性剖析，揭示了人工智能从“判别式感知”向“生成式理解”演进的深层逻辑。表 2 总结了具有代表性的 GUUM 模型及其技术架构、关键军事能力、实战应用场景以及主要安全风险。研究表明，GUUM 并非简单的战术工具叠加，而是驱动战争形态从“信息化”向“智能化”跃迁的核心变量。其对于现代战争的重塑作用，体现为决策环路的非线性压缩、战略资源的重新定义以及人机认知边界的深刻重构。

8.1 逻辑重构：基于“统一表征”的 OODA 环质变

GUUM 的核心价值在于通过多模态统一表征，打破了物理域、信息域与认知域之间的壁垒，从而在根本上改变了 OODA 环的运行机理。

1) 从“信息累积”到“认知涌现”。传统的 ISR 系统依赖海量数据的线性堆叠，而 GUUM 通过 RAG 与跨模态对齐技术，实现了异构数据的语义级融合。这种“多模态通感”能力使得机器能够在观察阶段即完成对战场态势的深层理解与补全，将情报处理从单纯的“去噪”提升为“基于概率的认知重构”，极大地降低了战场迷雾带来的信息熵。

2) 从“串行传导”到“并行博弈”。在决策与行动环节，GUUM 的上下文学习与 VLA 架构使得“判断-决策-行动”不再是离散的时间切片。模型能够在推理过程中同时进行意图解构、方案生成与动作规划，这种“感知-行动”的同构性使得 OODA 环呈现出高频、并行的网状特征。这不仅在物理时间上实现了对敌方决策周期的“降维打击”，更在认知空间内构建了具备自适应与反事实推理能力的博弈优势。

8.2 战略重定义：智能化战争的三大基石

随着 GUUM 介入作战全流程，支撑战争胜负的底层逻辑已发生转移，数据、算力与信任构成了新的战略三角。

1) 数据即新型高能弹药。数据的价值不再取决于规模，而取决于其“模态转换与指令微调”

表 2 不同架构的 GUUM 军事能力与安全风险
Table 2 Military capabilities and security risks of different GUUM architectures

模型/系统名称	核心技术架构	关键军事能力	实战应用场景	主要安全风险
COA-GPT ^[26]	LLM+In-Context Learning	自动化条令对齐 冲突检测 方案生成	战役规划 快速参谋作业 红蓝推演	幻觉导致方案不可行,过度 依赖导致人类思维退化
RT-2 ^[88] / OpenVLA ^[86]	Vision-Language- Action Transformer	语义-动作迁移 链式思维推理 零样本泛化	战场排爆 物资搬运 无人机自主控制	FreezeVLA ^[85] 对抗性攻击, 物理环境适应性差
Maven Smart System ^[24]	Computer Vision+ LLM+RAG	多源情报融合 自动化目标识别 交互式分析	全域态势感知 目标筛选 情报简报生成	数据投毒 情报来源混淆 隐私泄露
Lumina- DiMOO ^[84]	Discrete Diffusion	并行全域感知 跨模态语义对齐 感知-生成同步	跨域协同通信 综合态势实时生成与 预测	认知迷雾 算力瓶颈
Unified-IO 2 ^[17] /NExT- GPT ^[18]	Any-to-Any Multimodal Transformer	任意模态互转 统一语义理解	跨军种通信 语音转图 综合态势显示	模型庞大难以边缘部署 多模态对齐误差
Generative SAR ^[35]	Denoising Diffusion Probabilistic Models	超分辨率重建 图像翻译 去噪增强	全天候侦察 隐蔽目标发现 训练数据生成	生成虚假纹理细节(幻觉) 计算延迟高

的质量，高质量的、经过战术语义对齐的异构数据是训练军事 AI 世界模型的核心燃料。谁掌握了更符合真实物理规律与作战条令的数据集，谁就能训练出具备更高“战术智商”的模型，从而在认知对抗中占据制高点。

2) 算力即弹性战略纵深。GUUM 对推理资源的巨大消耗，使得算力网络成为新的关键基础设施。未来的战略纵深不再仅仅是地理空间的广度，而是算力拓扑的韧性。构建支持云边协同、抗毁重构的分布式算力网络，如美军联合作战云能力 (joint warfighting cloud capability, JWCC) 计划，是在高烈度电子战与物理打击下维持指挥控制体系存活与运转的物理基础。

3) 信任即认识论防线。生成式 AI 的幻觉问题与深伪攻击将战争的防线推向了认识论安全的层面。当战场信息的可信度被算法系统性稀释时，建立基于 C2PA 的密码学溯源机制，以及针对 VLA 模型的对抗性防御体系，不再是单纯的技术补丁，而是确立指挥官决策合法性与维护指挥链条稳定性的最后一道防线。

8.3 展望：在“不可能三角”中的动态平衡

尽管 GUUM 展现了巨大的变革潜力，但其未来的实战化部署仍需突破由“泛化能力、实时响应与可解释性”构成的“不可能三角”制约。

1) 泛化与实时的矛盾。追求极致的战场泛化能力往往意味着巨大的参数规模与推理延迟。未

来的研究需探索“快慢系统”的混合架构，即利用小规模 VLA 模型实现毫秒级的直觉式反应（快系统），同时调用云端超大模型进行深度的战术逻辑推理（慢系统）。

2) “黑盒”与信任的张力。面对神经网络的不可解释性，单纯的“黑盒”模型难以承载致死性武器的道德重负。神经符号人工智能将是一个关键突破口，通过将符号逻辑的严谨性嵌入深度学习的概率空间，有望在保留生成式 AI 创造力的同时，赋予其行为可验证、可追溯的逻辑约束。

综上所述，生成理解统一模型正在重写现代战争的源代码。未来的军事优势将属于那些能够率先解决上述矛盾，在“算法算力”与“人类智慧”之间建立完美人机共生关系的军队。

参考文献

[1] 郑逸宁，余镇，李不凡，等. 大语言模型的工具使用综述[J]. 自动化学报，2025，51(11)：2371-2386.
Zheng Yining, Yu Zhen, Li Bufan, et al. Survey of tool use in large language models[J]. Acta Automatica Sinica, 2025, 51(11): 2371-2386.
[2] 马翌硕，张光南，刘亚婷，等. 视觉-语言模型研究综述[J/OL]. 计算机技术与发展，2026，36(3)：1-10.
Ma Yishuo, Zhang Guangnan, Liu Yating, et al. A survey of vision-language models[J/OL]. Com-

- puter Technology and Development, 2026, 36(3): 1-10.
- [3] 刘紫权, 史旭阳, 李珂, 等. 视觉语言模型架构发展综述[J/OL]. 计算机应用, (2025-08-29)[2025-12-01]. <https://link.cnki.net/urlid/51.1307.TP.20250828.1345.002>.
- Liu Ziquan, Shi Xuyang, Li Ke, et al. Review of vision-language model architecture develop[J/OL]. Journal of Computer Applications, (2025-08-29)[2025-12-01]. <https://link.cnki.net/urlid/51.1307.TP.20250828.1345.002>.
- [4] 张熙, 杨小汕, 徐常胜. ChatGPT 及生成式人工智能现状及未来发展方向[J]. 中国科学基金, 2023, 37(5): 743-750.
- Zhang Xi, Yang Xiaoshan, Xu Changsheng. Current state and future development directions of ChatGPT and generative artificial intelligence[J]. Bulletin of National Natural Science Foundation of China, 2023, 37(5): 743-750.
- [5] 林国军, 龚涛, 林通. 图像分类任务中的注意力机制研究综述[J]. 电讯技术, 2025, 65(8): 1349-1362.
- Lin Guojun, Gong Tao, Lin Tong. A survey of attention mechanisms in image classification tasks[J]. Telecommunication Engineering, 2025, 65(8): 1349-1362.
- [6] 王永强, 徐林峰, 陈炜. 多模态数据下的目标检测与识别研究综述[J]. 工业控制计算机, 2025, 38(10): 90-91, 94.
- Wang Yongqiang, Xu Linfeng, Chen Wei. A review of object detection and recognition research with multimodal data[J]. Industrial Control Computer, 2025, 38(10): 90-91, 94.
- [7] Fang Li, Yang Zixian, Yang Chen. Radar signal sorting technology based on image processing and hough transform[C]//International Conference on Microwave and Millimeter Wave Technology, 2018: 1-3.
- [8] Sue E K, Chou P H, Tomer K, et al. The future of collaborative human-artificial intelligence decision-making for mission planning[J]. Frontiers in Psychology, 2022, 13: 850628.
- [9] Agarwala N. Robots and artificial intelligence in the military[J]. Obrana a Strategie, 2023, 23(2): 83-100.
- [10] 侯立志, 李立伟, 陈鑫, 等. 智能化作战体系[J]. 兵工自动化, 2025, 44(7): 66-69.
- Hou Lizhi, Li Liwei, Chen Xin, et al. Intelligent combat system[J]. Ordnance Industry Automation, 2025, 44(7): 66-69.
- [11] 白成超, 张琦, 谢旭东, 等. 面向复杂决策的 OO-DA 环: 智能赋能与认知增强[J]. 指挥与控制学报, 2024, 10(3): 284-297.
- Bai Chengchao, Zhang Qi, Xie Xudong, et al. OO-DA ring theory for complex decision-making: intelligent empowerment and cognitive enhancement[J]. Journal of Command and Control, 2024, 10(3): 284-297.
- [12] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.
- [13] Zhang Zhijun, Zhang Jian, Zhang Xiaodong, et al. A comprehensive overview of generative AI (GAI): technologies, applications, and challenges[J]. Neurocomputing, 2025, 632: 129645.
- [14] Cao Yihan, Li Siyu, Liu Yixin, et al. A survey of AI-generated content (AIGC)[J]. ACM Computing Surveys, 2025, 57(5): 1-38.
- [15] Yuan Yuan, Li Zhaojian, Zhao Bin, et al. A survey of multimodal learning: methods, applications, and future[J]. ACM Computing Surveys, 2025, 57(7): 1-34.
- [16] Zhou Fan, Chen Yuyi, Wen Jun, et al. Episodic task agnostic contrastive training for multi-task learning[J]. Neural Networks, 2023, 162: 34-45.
- [17] Zhao Shanshan, Zhang Xinjie, Guo Jintao, et al. Unified multimodal understanding and generation models: advances, challenges, and opportunities[EB/OL]. [2025-12-01]. <https://doi.org/10.48550/arXiv.2505.02567>.
- [18] Lu Jiasen, Clark C, Lee S, et al. Unified-IO₂: scaling autoregressive multimodal models with vision, language, audio, and action[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2024: 26429-26445.
- [19] Wu Shengqiong, Fei Hao, Qu Leigang, et al. NE-xT-GPT: any-to-any multimodal LLM[EB/OL]. (2024-06-25)[2025-12-01]. <https://doi.org/10.48550/arXiv.2309.05519>.
- [20] Li Shufan, Katsamakas K, Gokul A, et al. OmniFlow: any-to-any generation with multi-modal rectified flows[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2025: 13178-13188.
- [21] Kirichenko D. Artificial intelligence's growing role in modern warfare[EB/OL]. (2025-08-21). [2025-12-01]. <https://warroom.armywarcollege.edu/articles/ai-growing-role>.
- [22] Clark B, Patt D, Schramm H. Mosaic warfare:

- exploiting artificial intelligence and autonomous systems to implement decision-centric operations[EB/OL]. (2020-02-11) [2025-12-01]. https://csbaonline.org/uploads/documents/Mosaic_Warfare.pdf.
- [23] Deptula D A, Penney H. Mosaic warfare[J/OL]. (2019-11-01) [2025-12-01]. <https://www.airandspaceforces.com/article/mosaic-warfare>.
- [24] Kevin J S, Duska J. Project maven: how AI quietly entered the kill chain[EB/OL]. (2025-03-23) [2025-12-01]. <https://aiweapons.tech/projectmaven-how-ai-quietly-entered-the-kill-chain>.
- [25] Dong Qingxu, Li Lei, Dai Damai, et al. A survey on in-context learning [C]//Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. Miami, Florida, USA: Association for Computational Linguistics, 2024: 1107-1128.
- [26] Goecks V G, Waytowich N. COA-GPT: generative pre-trained transformers for accelerated course of action development in military operations[C]//2024 International Conference on Military Communication and Information Systems (ICMCIS). Piscataway, NJ, USA: IEEE, 2024: 1-10.
- [27] Brohan A, Brown N, Carbajal J, et al. RT-2: vision-language-action models transfer web knowledge to robotic control[EB/OL]. (2023-07-28) [2025-12-01]. <https://doi.org/10.48550/arXiv.2307.15818>.
- [28] 张龙, 王数, 雷震, 等. AIGC军事大模型评估体系框架研究[J]. 战术导弹技术, 2025(1): 42-52.
Zhang Long, Wang Shu, Lei Zhen, et al. Research on the framework of AIGC military large-model evaluation system[J]. Tactical Missile Technology, 2025(1): 42-52.
- [29] Wang Deyi, Zhang Jisheng, Zheng Shicheng. Multi-source information data fusion method under complex battlefield situation[J]. Journal of Physics: Conference Series, 2020, 1631(1): 012113.
- [30] Beese J, Aier S, Haki K, et al. The impact of enterprise architecture management on information systems architecture complexity[J]. European Journal of Information Systems, 2023, 32(6): 1070-1090.
- [31] Lu Qiang, Sun Xia, Gao Zhizezhang, et al. Co-ordinated-joint translation fusion framework with sentiment-interactive graph convolutional networks for multimodal sentiment analysis[J]. Information Processing & Management, 2024, 61(1): 103538.
- [32] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. (2021-06-03) [2025-12-01]. <https://doi.org/10.48550/arXiv.2010.11929>.
- [33] Girdhar R, El-Nouby A, Liu Zhuang, et al. ImageBind: one embedding space to bind them all [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2023: 15180-15190.
- [34] Sahoo P, Singh A K, Saha S, et al. A systematic survey of prompt engineering in large language models: techniques and applications[EB/OL]. (2025-03-16) [2025-12-01]. <https://doi.org/10.48550/arXiv.2402.07927>.
- [35] Shahrezaei H I, Kim H C. Resolutional analysis of multiplicative high-frequency speckle noise based on SAR spatial de-speckling filter implementation and selection[J]. Remote Sensing, 2019, 11(9): 1041.
- [36] Perera M V, Nair N G, Bandara W G C, et al. SAR despeckling using a denoising diffusion probabilistic model[J]. IEEE Geoscience and Remote Sensing Letters, 2023, 20: 4005305.
- [37] Chen Zilong, Zhang Caiguang, Wan Chengyu, et al. DADSR: degradation-aware diffusion super-resolution model for object-level SAR image[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2025, 18: 19496-19510.
- [38] Chen Zilong, Zhang Siqian, Xiong Boli. Super-resolving SAR images with diffusion models: a dual evaluation of metrics and applications[C]//2024 IEEE 17th International Conference on Signal Processing (ICSP). Piscataway, NJ, USA: IEEE, 2025: 340-346.
- [39] Liu Yang, Chen Weixin, Bai Yongjie, et al. Aligning cyber space with physical world: a comprehensive survey on embodied AI[J]. IEEE/ASME Transactions on Mechatronics, 2025, 30 (6): 7253-7274.
- [40] Din M U, Akram W, Saoud S L, et al. Vision language action models in robotic manipulation: a systematic review[EB/OL]. (2025-07-14) [2025-12-01]. <https://doi.org/10.48550/arXiv.2507.10672>.
- [41] Driess D, Xia Fei, Mehdi S M S, et al. PaLM-E: an embodied multimodal language model[EB/OL]. (2023-03-06) [2025-12-01]. <https://doi.org/10.48550/arXiv.2303.03378>.
- [42] MARR J J. The military decision making process: making better decisions versus making decisions better[R/OL]. (2001-01-01) [2025-12-01]. <https://api.semanticscholar.org/CorpusID: 108613720>.
- [43] Hoijsink M, Hardeveld P A. Machine learning and the platformization of the military: a study of Google's machine learning platform TensorFlow[J].

- International Political Sociology, 2022, 16 (2) : olab036.
- [44] Palantir Technologies. Palantir artificial intelligence platform (AIP) [EB / OL]. [2025-12-01]. <https://www.palantir.com/platforms/aip/>.
- [45] Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York, USA: ACM, 2020: 9459-9474.
- [46] Liggins M, Hall D, Llinas J. Handbook of Multi-Sensor Data Fusion: Theory and Practice[M]. Boca Raton: CRC Press, 2008: 732-771.
- [47] JIEYAB89. OSINT Cheat Sheet [EB/OL]. GitHub. (2025-08-21) [2025-12-01]. <https://jiejab89-os-int.gitbook.io/jiejab89-osint-cheat-sheet-wiki-tips>.
- [48] Zhang Chudi, Wang Lei, Jiang Rundong, et al. Radar jamming decision-making in cognitive electronic warfare: a review[J]. IEEE Sensors Journal, 2023, 23(11): 11383-11403.
- [49] Zhang Chudi, Song Yunqi, Jiang Rundong, et al. A cognitive electronic jamming decision-making method based on Q-learning and ant colony fusion algorithm[J]. Remote Sensing, 2023, 15(12): 3108.
- [50] Kaasaragadda Y, Kokalj-Filipovic S. ReFormer: generating radio fakes for data augmentation[EB/OL]. (2024-12-31) [2025-12-01]. <https://doi.org/10.48550/arXiv.2501.00282>.
- [51] Carmack J, Bhatia A, Robinson J, et al. Neural network generative models for radio frequency data [C]//2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). Piscataway, NJ, USA: IEEE, 2021: 577-582.
- [52] Zhang Chudi, Yang Biao, Wang Lei, et al. A cognitive jamming decision-making method based on heuristic improved A2C algorithm[J]. IEEE Transactions on Vehicular Technology, 2025, 74 (2): 2871-2883.
- [53] Zhang Chudi, Yang Biao, Ji Wenshuai, et al. Cognitive jamming policy generation based on A2C algorithm[C]//2024 International Radar Symposium (IRS). Piscataway, NJ, USA: IEEE, 2024: 33-38.
- [54] Salvato E, Fenu G, Medvet E, et al. Crossing the reality gap: a survey on sim-to-real transferability of robot controllers in reinforcement learning [J]. IEEE Access, 2021, 9: 153171-153187.
- [55] Xian Zhou, Gervet T, Xu Zhenjia, et al. Towards generalist robots: a promising paradigm via generative simulation[EB / OL]. (2023-08-30) [2025-12-01]. <https://doi.org/10.48550/arXiv.2305.10455>.
- [56] Vinyals O, Babuschkin I, Czarnecki W M, et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning[J]. Nature, 2019, 575(7782): 350-354.
- [57] Madison A, McDowell K, Goecks V G, et al. New challenges for future C2: commanding soldier-machine partnerships[EB/OL]. (2025-03-11) [2025-12-01]. <https://doi.org/10.48550/arXiv.2503.08844>.
- [58] Madison A, Novoseller E, Goecks V G, et al. Scalable interactive machine learning for future command and control[C]//2024 International Conference on Military Communication and Information Systems (ICMCIS). Piscataway, NJ, USA: IEEE, 2024: 1-10.
- [59] Mayer M. Trusting machine intelligence: artificial intelligence and human-autonomy teaming in military operations[J]. Defense & Security Analysis, 2023, 39(4): 521-538.
- [60] Gao Kun, Wang Hao, Nazarko J, et al. Adaptive decision method in C3I system[J]. Computational Intelligence and Neuroscience, 2022(1): 6967223.
- [61] Bernatskyi A, Kovalenko O. Adaptive control of the combat environment based on space-time entropy[J]. Communication Informatization and Cybersecurity Systems and Technologies, 2025, 1(7): 1-10.
- [62] Xu Jianhong, Liang Gongqian. Enhanced Q learning and deep reinforcement learning for unmanned combat intelligence planning in adversarial environments[J]. Scientific Reports, 2025, 15: 28364.
- [63] Varley J, Singh S, Jain D, et al. Embodied AI with two arms: zero-shot learning, safety and modularity[C]//2024 IEEE / RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway, NJ, USA: IEEE, 2024: 3651-3657.
- [64] Shao Hao, Qian Shengju, Xiao Han, et al. Visual CoT: advancing multi-modal language models with a comprehensive dataset and benchmark for chain-of-thought reasoning[EB/OL]. (2024-11-04) [2025-12-01]. <https://doi.org/10.48550/arXiv.2403.16999>.
- [65] Jond H B. A minimal model for emergent collective behaviors in autonomous robotic multi-agent systems[J]. IEEE Transactions on Cognitive and Developmental Systems, 2026, 18(2): 373-384.
- [66] 新华社研究院. 思想殖民——美国认知战的手段、根源及国际危害[R/OL]. [2025-12-01]. https://www.xhinst.net/jyyjqbsj/detail/20250904/19447831_全文_思想殖民美国认知战的手段根源及国际危害_1.html.

- [67] Yang Mingzhe, Wang Zhipeng, Liu Kaiwei, et al. Finding emergence in data by maximizing effective information[J]. *National Science Review*, 2024, 12(1): nwae279.
- [68] Mirsky Y, Lee W. The creation and detection of deepfakes: a survey[J]. *ACM Computing Surveys*, 2021, 54(1): 1-41.
- [69] Zhao Wenshuai, Queralta J P, Westerlund T. Sim-to-real transfer in deep reinforcement learning for robotics: a survey[C]//2020 IEEE Symposium Series on Computational Intelligence (SSCI). Piscataway, NJ, USA: IEEE, 2021: 737-744.
- [70] Kim D, Lee H, Cha J, et al. Bridging the reality gap: analyzing sim-to-real transfer techniques for reinforcement learning in humanoid bipedal locomotion[J]. *IEEE Robotics & Automation Magazine*, 2025, 32(1): 49-58.
- [71] Liu Nanjun, Cai Yinghao, Lu Tao, et al. Real sim real transfer for real-world robot control policy learning with deep reinforcement learning[J]. *Applied Sciences*, 2020, 10(5): 1555.
- [72] Wu Zhonghong, Pan Li, Yu Minggang, et al. A game-based approach for designing a collaborative evolution mechanism for unmanned swarms on community networks[J]. *Scientific Reports*, 2022, 12: 18892.
- [73] Bu Yajun, Yan Ye, Yang Yueneng. Advancement challenges in UAV swarm formation control: a comprehensive review[J]. *Drones*, 2024, 8(7): 320.
- [74] Marcelino W, Welch J, Clayton B, et al. Acquiring generative artificial intelligence to improve U. S. Department of Defense influence activities[EB/OL]. (2025-07-22)[2025-12-01]. https://www.rand.org/pubs/research_reports/RRA3157-1.html.
- [75] Qing Yun. Vision transformers (vits) for feature extraction and classification of AI-generated visual Designs[J]. *ISSS Journal of Micro and Smart Systems*, 2025, 13: 69459-69477.
- [76] Petráš Z. Conceptual approach to multi-domain operations[J]. *Vojenské Rozhledy*, 2023, 32(4): 66-85.
- [77] Révay M, Liska M. OODA loop in command & control systems[C]//2017 Communication and Information Technologies (KIT). Piscataway, NJ, USA: IEEE, 2017: 1-4.
- [78] Hagos D H, Rawat D B. Neuro-symbolic AI for military applications[J]. *IEEE Transactions on Artificial Intelligence*, 2024, 5(12): 6012-6026.
- [79] Amore L, Campolo A, Jacobsen B, et al. A world model: on the political logics of generative AI[J]. *Political Geography*, 2024, 113: 103134.
- [80] Xin Yi, Qin Qi, Luo Siqu, et al. Lumina-DiMOO: an omni diffusion large language model for multi-modal generation and understanding[EB/OL]. (2025-10-07)[2025-12-01]. <https://doi.org/10.48550/arXiv.2510.06308>.
- [81] Christopher J K, Bartoldson B R, Ben-Nun T, et al. Speculative diffusion decoding: accelerating language generation through diffusion[C]//Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies, 2025, 1: 12042-12059.
- [82] Aionesei G A, Panait C. Understanding multi-domain operations from the air force perspective[J]. *Review of the Air Force Academy*, 2024, 22(1): 61-70.
- [83] Kozlov V, Slyusar V, Tverdokhlibov V, et al. Integration and coordination of electronic warfare assets through largescale language models[J]. *Technology Audit and Production Reserves*, 2025, 1(2(81)): 54-61.
- [84] Wang Shafei, Bao Yanfei, Li Yan. The architecture and technology of cognitive electronic warfare [J]. *Scientia Sinica Informationis*, 2018, 48(12): 1603-1613.
- [85] Johnson J. Automating the OODA loop in the age of intelligent machines: reaffirming the role of humans in command-and-control decision-making in the digital age[J]. *Defence Studies*, 2023, 23(1): 43-67.
- [86] Wang Xin, Li Jie, Wen Zejia, et al. FreezeVLA: action-freezing attacks against vision-language-action models[EB/OL]. (2025-09-24)[2025-12-01]. <https://doi.org/10.48550/arXiv.2509.19870>.
- [87] Kim M J, Pertsch K, Karamcheti S, et al. OpenVLA: an open-source vision-language-action model [EB/OL]. (2024-09-05)[2025-12-01]. <https://doi.org/10.48550/arXiv.2406.09246>.
- [88] Lu Hui, Yu Yi, Yang Yiming, et al. When robots obey the patch: universal transferable patch attacks on vision-language-action model[EB/OL]. (2025-11-30)[2025-12-01]. <https://doi.org/10.48550/arXiv.2511.21192>.
- [89] Li Ke, Wang Di, Zhu Wenxuan, et al. Physical adversarial patch attack for optical fine-grained aircraft recognition[J]. *IEEE Transactions on Information Forensics and Security*, 2025, 20: 436-448.
- [90] Sun Bo, Tsai N H, Liu Fangchen, et al. Adversarial defense by stratified convolutional sparse cod-

- ing[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 11439–11448.
- [91] Hu Aodi, Zhang Zhiyong, Quan Gaoyuan, et al. IDPA: indiscriminate data poisoning attacks targeting pre-trained encoder based on contrastive learning[J]. *ACM Transactions on Privacy and Security*, 2025, 28(4): 1–22.
- [92] Seddik M E A, Chen S W, Hayou S, et al. How bad is training on synthetic data? A statistical analysis of language model collapse[EB/OL]. (2024-04-07) [2025-12-01]. <https://doi.org/10.48550/arXiv.2404.05090>.
- [93] Shumaliyov I, Shumaylov Z, Zhao Yiren, et al. AI models collapse when trained on recursively generated data[J]. *Nature*, 2024, 631(8022): 755–759.
- [94] Sahoo P, Meharia P, Ghosh A, et al. A comprehensive survey of hallucination in large language, image, video and audio foundation models[C]//Findings of the Association for Computational Linguistics: EMNLP 2024. Miami, Florida, USA: Association for Computational Linguistics, 2024: 11709–11724.

Review of the Applications of Generative Understanding Unified Models in Military Command and Control

LI Min¹, YAN Song^{1, 2}, ZHANG Yusen¹, LI Ruixuan¹,
GOU Yao³, HE Yujie^{1*}

(1. Rocket Force University of Engineering, Xi'an 710025, China;

2. NEL-BITA, University of Science and Technology of China, Hefei 230026, China;

3. Institute of Military Intelligence, Academy of Military Sciences, Beijing 100091, China)

ABSTRACT: As artificial intelligence technology transitions from single-modal discriminative paradigms to multi-modal generative paradigms, the generative understanding unified model (GUUM) has emerged as a novel technical architecture that reshapes the underlying logic of modern military power. First, this study systematically analyzed the construction mechanisms of GUUM, revealing that its general capabilities stemmed from the unified modeling and analysis of multi-modal data. It further examined the distinctions in its application at the perception level. Finally, the review focused on interactions within the physical domain, analyzing how generative planning and diffusion policy-based embodied intelligence transformed the physical battlefield. At the application level, the review incorporated specific case studies to provide an in-depth exploration of GUUM efficacy in the automated course of action generation and wargames, the evolution of intelligence fusion utilizing GUUM, diffusion model enhancement techniques for synthetic aperture radar images, and the tactical value of GUUM in cognitive and electronic warfare spectrum generation. Drawing on the latest research on adversarial attacks, this review exposed the vulnerabilities of unified models when facing physical perturbations, data poisoning, and model hallucinations. This study aimed to provide a comprehensive assessment for military decision-makers and defense researchers, demonstrating how GUUM served as a key enabling technology for advancing novel operational concepts from scenarios to reality and facilitating all-domain command and control while offering defensive recommendations to address security risks.

KEYWORDS: generative understanding unified model; vision-language-action model; command and control; cognitive warfare

(编辑: 吴文冬)