

特色专题

面向无人系统的具身社会感知智能: 流程与框架

范丽丽¹, 史光宇², 陈恺舸¹, 李创³, 曾昌贤⁴, 邓方^{1*}

摘要 自主智能无人系统在动态复杂、多主体耦合、信息不完备且社会约束强的真实开放环境中运行时, 面临合规性建模不足、社会风险感知受限、协同冲突突出及异常响应滞后等关键挑战。为此, 提出一种具身社会感知智能框架, 该框架将具身感知(本体、内部、外部、交互和意图)与社会雷达融合, 并引入代理式人工智能(Agentic AI)作为顶层决策与控制机制, 实现多层级、自主化的认知决策。整体采用感知、推理、执行、反馈与元控制 5 层结构, 实现从多模态感知到合规行为生成的动态闭环。通过融合物理与社会环境信息, 该框架显著提升了自主智能无人系统在城市治理、应急救援和社会安全等复杂不确定场景中的任务适应性、群体协同效率与合规可靠性, 为实现可信赖、可解释、可持续的自主智能系统提供新的技术路径。

关键词 具身感知; 社会雷达; Agentic AI; 自主智能无人系统

在俄乌冲突这一智能化对抗的“试验场”上, 多种新技术与新装备不断涌现并投入使用, 无人系统与人工智能的深度融合不仅凸显了现代对抗的高科技特征, 也在重塑传统对抗场格局。双方先后部署了数十种无人装备, 涵盖无人机、无人车、无人舰艇等多种类型。乌克兰的“愤怒”“列雷卡”“惩罚者”“旁观者”等型号无人机具备情报、监视与侦察能力, 能够对对方部队位置和动向进行精确感知; 俄罗斯则主要依靠“奥兰-10”等中小型多用途无人机执行侦察监视任务。与此同时, 人工智能在信息与认知领域的作用同样突出: 交战双方广泛使用智能算法处理来自媒体与社交网络的开源内容, 用于识别、传播甚至反制虚假信息,

舆论与情绪信号由此直接嵌入博弈过程中^[1]。更具象征意义的是, 对抗场透明度已不再完全依赖前线侦察兵的数量。当一名乌克兰农民用手机拍摄对方车队并上传至 Diia 政务应用程序时, 这一“全民传感器”便与无人机、卫星等具身感知手段, 共同织就了前所未有的信息获取网络^[1]。在此基础上, 乌克兰依托 Delta 态势感知系统与 Palantir Meta Constellation, 将来自无人机、卫星、前沿传感器与盟友情报的多源数据整合进统一平台, 生成数字化对抗场模型^[2], 实时识别对方阵地并优化打击方案, 从而显著提升了指挥与决策的效率与精度^[1]。Meta Constellation 等工具也被用来聚合商业卫星与热传感无人机图像, 为目标位置预测与应对方案提供支持^[3]。

这种由多源信息汇聚、经智能算法解析并快速转化为行动方案的方式, 不仅限于国际对抗冲突, 也为民用领域提供启示。在大规模自然灾害或城市应急事件中, 救援力量往往因信息不对称而陷入被动局

1. 北京理工大学人工智能学院, 北京 100081

2. 北京理工大学(珠海)人工智能学院, 珠海 519088

3. 中国人民解放军军事科学院军事智能研究院, 北京 100091

4. 重庆大学机械与运载工程学院, 重庆 400044

收稿日期: 2025-09-06; 修回日期: 2025-10-11

作者简介: 范丽丽, 助理教授, 研究方向为智能无人系统感知, 电子信箱: lilifan@bit.edu.cn; 邓方(通信作者), 教授, 研究方向为群体智能、智能群系统、可穿戴泛在系统, 电子信箱: dengfang@bit.edu.cn

引用格式: 范丽丽, 史光宇, 陈恺舸, 等. 面向无人系统的具身社会感知智能: 流程与框架[J]. 科技导报, 2025, 43(20): 37-47; doi:10.3981/j.issn.1000-7857.2025.09.00071

面:前线救援人员依赖目视侦查,覆盖面有限;传统通信设施受损,情况难以及时上报;而灾情复杂多变,对救援速度和资源分配提出了极高要求。近年来,多个案例展现了这一设想的部分实现。例如,2017年,在飓风 Harvey 与 Irma 过后,美国南部的 2 个救援团队利用无人机进行灾后损毁评估,同时收集社交媒体、新闻报道等非正式信息源补充现场观察,快速判断哪些道路被毁、哪些区域受灾最严重,从而更有效地调度救援资源^[4]。在某次自然灾害中,组织与自愿者合作,通过自动化图像处理系统分析约 28 万张社交媒体图片,评估灾害造成的损失程度。经过领域专家对约 2.9 万张系统处理图像的反馈,发现该系统的准确率达到了 76%^[5]。此外,已有研究将无人机视觉与卫星遥感图像相结合,通过联合分析实现火源位置识别和道路通行条件评估,并在火情蔓延模型的基础上对逃生路线进行动态规划^[6]。然而,这些尝试大多仍停留在研究或局部实验阶段。2024 年,巴西 Rio Grande do Sul 州的洪灾中,一项研究发现,民用无人机的活动对官方搜索救援流程造成了显著干扰。即使当局设立了无人机禁飞区,但也影响了救援规划、地面队伍调度与通信组织^[7]。这些案例表明,尽管前景可期,但在统一规则、实时性、协同性与决策集成等方面仍存在明显不足。因此,探索新的理论框架与技术路径显得尤为迫切。

无论是在对抗环境中提升透明度,还是在灾害救援中加快响应速度,都表明一个共同趋势:对物理环境的实时感知、对社会因素的动态反馈以及基于智能算法的自主决策正在逐步融合。未来,智能无人系统的发展不应再局限于单一的感知或局部的功能提升,而是需要在任务理解、决策闭环和多主体协同等方面形成整体能力。换句话说,将具身感知、社会信息获取与智能体决策机制有机结合,不仅是应对复杂现实场景的必然选择,也将成为推动自主智能无人系统走向可靠、可控和工程化应用的关键路径。

由此可见,在复杂开放环境中,单纯依靠物理传感已难以支撑无人系统的可靠运行。如何同时吸纳社会层面的规则、事件和群体反馈,并与自主智能体的决策过程深度耦合,成为未来发展的关键。基于此,提出“具身社会感知智能”这一概念,并构建了面向无人系统的流程与框架,力图在智能城市治理、紧

急响应与灾害管理、智能制造与工业优化等典型场景中,为无人系统提供更加安全、高效、易理解的解决方案。

1 从具身感知到具身社会感知智能

从具身感知到具身社会感知智能的转变,不仅拓展了感知能力,也是实现更高级别智能的关键。传统具身感知使无人系统能够精准感知物理环境并评估自身状态,但在复杂且动态变化的社会环境中,单一的物理感知已无法满足需求。因此,将社会感知信息纳入决策框架成为关键发展方向。具身社会感知智能通过融合物理感知与社会反馈,赋予系统更全面的环境理解和决策能力,使其不仅能够响应物理世界,还能理解并应对社会行为和群体动态。这种融合提升了无人系统的认知和决策能力,使其在应对物理环境变化的同时,也能智能响应社会情境中的动态变化。以下将探讨具身感知、社会感知和智能决策如何协同工作,推动无人系统在各领域的深度应用与发展。

1.1 具身感知

具身感知是自主智能无人系统实现环境理解和任务执行的核心基础,涵盖了对自身状态、外部环境以及人机交互的全面感知(图 1)。近年来,随着多模态传感与融合技术的快速发展,具身感知在复杂环境建模、运动预测和人机交互等方面展现出巨大潜力。具身感知并非单一模态的叠加,而是一个覆盖自身、交互、环境、健康与社会意图的全域感知立体体系。其一,本体感知使系统能够实时感知自身的位置、姿态和受力状态。通过旋转/线性编码器和惯性测量单元,系统可以实现位置、速度和加速度的实时估计。多轴力/力矩传感器则用于监测关节载荷,支撑高精度的控制与碰撞检测。近年来,研究人员进一步在结构中嵌入光纤布拉格光栅与柔性应变片,以捕捉微小形变与载荷的捕捉,使机器人像生物关节一样具备敏锐的“本体感觉”^[8]。其二,交互感知赋予系统“皮肤”,通过柔性电子皮肤的分布式节点,系统能够同时感知压力、剪切力与温湿度。触觉阵列则解析摩擦力与纹理信息,从而实现精细抓取与稳定接触。更前沿的研究将视觉与触觉统一建模,使机器人在“看”的同时也能

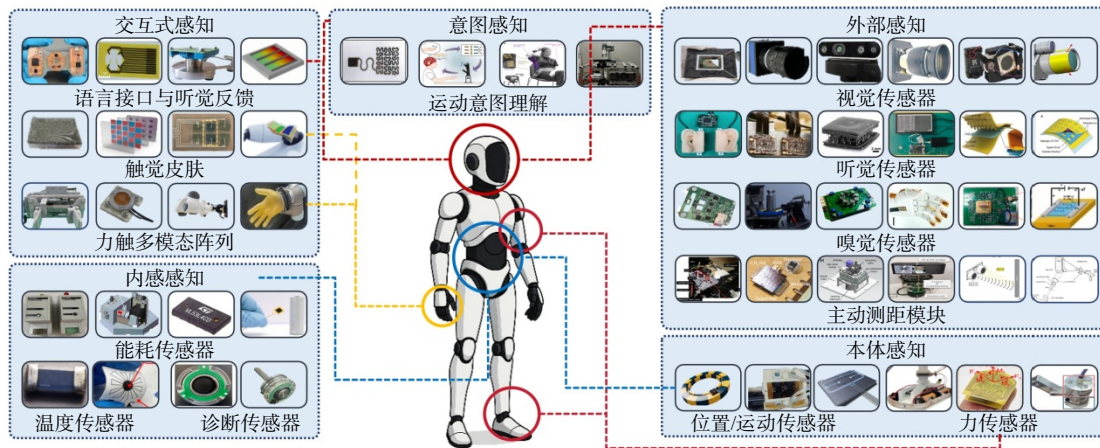


图1 5类感知常见传感器

“摸”,从而实现跨模态的接触理解^[9]。其三,外界感知是无人系统连接环境的主要通道。彩色相机、深度相机、事件相机和热成像相机分别提供颜色、几何、动态和低光条件下的信息;声学传感器用于声源定位与识别^[10];仿生嗅觉模块可以检测气体成分。结合激光雷达、毫米波雷达和超声波传感器,系统可以构建三维语义地图,从而实现对复杂环境的整体理解^[11]。其四,内部感知聚焦于“身体健康”。能耗传感器监测电机电流与电压,温度传感器关注电池和高功率部件的热状态^[12];嵌入材料中的应变计或加速度计帮助发现轴承、齿轮等关键部件的裂纹与磨损,从而实现在线健康监测与预测性维护^[13]。其五,意图感知扩展了具身感知的社会性边界。柔性表情传感器可实时识别微表情,脑电与肌电信号解码技术则帮助理解人类的操作计划与动作意图^[14]。同时,基于Wi-Fi、毫米波雷达与红外阵列的非接触式方法^[15]正在用于识别人群流线与社交行为特征,这在群体交互与协作任务中尤显重要。

本体感知、交互感知、外界感知、内感知和意图感知这5类感知共同支撑了智能无人系统在复杂环境中的稳健运行,它们不仅使机器具备“自知”能力,也使其能够感知环境和识别健康状态,甚至初步具备理解人类意图的潜力。然而,这些感知大多仍停留在物理和个体层面,难以触及社会语境与群体协同的复杂性。其根本原因在于,尽管传感器与融合算法能够将世界还原为几何、纹理、速度与受力等物理量,但它们难以回答诸如“此刻的行动是否可行”“何者应优先”“此行动是否引发社会风险”等更高层次的

问题。例如,在城市应急场景中,急救无人车可依赖激光雷达与相机准确识别道路环境,但无法提前获知前方道路因临时交通管制而封闭。结果是车辆到现场才被迫绕行,延误救援。这一情形体现了具身感知的局限,尽管感知能力强大,却无法捕捉社会信号。在对抗侦察场景中,尽管无人侦察机能够精确探测地形,但是难以判断临时设立的安全停火区或人道主义通道是否有效,从而可能导致违规进入并引发国际批评。为此,联合国(含人权理事会相关机制)在2024年的文件与讨论中,主张对“自主武器系统”建立以“保持人类控制”为核心、符合国际法/人权法的规范与合规框架^[16]。民用救援领域也有类似情况发生:在欧洲某地的森林火灾中,无人机尽管提供了高分辨率的热成像图像,却未能结合社交媒体上关于火势蔓延和人员被困的实时信息,导致救援调度延误,进而扩大了损失。

这些案例揭示了单纯具身感知的4大局限:(1)任务起始和优先级判断混乱,缺乏对“紧急”与“重要”的正确判别;(2)合规性不足,难以将法规、礼仪与临时管制转化为可计算的约束条件;(3)社会风险感知不足,群体拥挤、情绪波动与伦理敏感点常在感知缺失下爆发;(4)突发响应滞后,系统只能被动等待物理信号,而无法基于社会先兆提前调整。

1.2 社会雷达

1.2.1 社会雷达的起源

社会雷达的概念并非凭空产生,而是历史经验、学理积淀与现实事件相互作用的结果。早在中国古代,《诗经》中的“风、雅、颂”体系与相传的采诗制度

相联系。据说周王曾设采诗之官,采集诸侯与民间歌谣,以察民情、观政得失与社会风貌。此种通过信息采集与反馈来感知社会并调整治理的方式,与当今依托社交媒体与网络动态追踪群体态势的“社会雷达”在原理上相通,可视为其思想雏形。

在学理层面,政治学家 Lasswell 在 20 世纪中期提出,社会需要具备环境监测功能,以持续追踪公众意见与社会态势。传播学者 Schramm 在 20 世纪 80 年代借用“雷达”这一比喻,首次提出“社会雷达”一词,用以形象说明社会系统如何通过采集、识别与反馈信息,像物理雷达一样调整自身运行^[17]。这一表述为社会雷达提供了理论框架。

现实事件则推动了社会雷达快速走向实践。2011 年,“阿拉伯之春”成为社会雷达走入公众视野的重要契机。在突尼斯、埃及等国家,推特上的标签运动与脸书上的集会呼吁,使分散的社会情绪在极短时间内迅速聚合并转化为群体行动^[18]。Mathieu 等在《IEEE Intelligent Systems》撰文指出,这场运动“让社交媒体成为新的‘社会雷达’,它不仅能捕捉社会情绪,还能触发社会变革”^[19]。进入大数据与人工智能时代,社会雷达逐渐演变为一种面向公共治理和网络安全智能服务平台。通过关键词扩展、情感分析与事件检测等方法,社会雷达可以实时追踪社交媒体动态,生成社会态势画像,为政府和企业提供风险预警与决策支持。由此,社会雷达完成了从采风式的经验性探索,到社会态势感知的理论化设想,再到智能平台的工程化实现的发展过程,逐步成长为理解社会动态与群体行为的重要工具。

1.2.2 社会雷达的发展

社会雷达的研究正逐步从概念探索走向体系化发展。早期工作主要停留在社会科学和传播学领域,强调其类似物理雷达的功能:通过收集、识别和反馈社会信号,帮助组织理解公众态势并调整行为^[20]。如今,研究开始聚焦于如何将社交媒体、在线论坛和新闻报道中的多源信息转化为可计算的社会态势图。例如,Mathieu 等^[21]提出了基于关键词扩展与情感分析的社会雷达 workflow,用于对抗与安全场景的风险监测。

近年来,社会雷达的研究进一步扩展到网络安全、交通、城市治理和智能车辆等领域。Fan 等^[22]指出,社会雷达已成为网络空间中识别潜在威胁的重要

工具,能够通过情感分析与趋势建模预测网络信息演变,并为政府和企业提供风险预警与决策支持。在交通领域,Cao 等^[23]提出交通社会雷达概念,通过收集公众对交通状况的意见与情绪,实现智能交通系统的人性化感知与优化。同时,智能车辆领域也引入社会雷达的框架,使车辆在物理感知的基础上,能够通过社会雷达理解人群行为与社会语境,提升群体交互与安全性。最新研究将多普勒效应引入社会雷达,提出社会多普勒雷达的概念,用以刻画社会情绪和群体行为的动态变化速率与方向。这一方法借鉴了物理雷达中利用多普勒频移表征目标径向速度的原理,将信息流和情感信号的频移映射为社会态势的演化趋势,从而在风险预警和群体行为预测中提供更高的敏感性和前瞻性。图 2 展示了物理雷达信号处理流程与社会雷达信息感知阶段之间的对应关系,物理空间中的雷达与网络空间中的社会雷达都依赖反馈捕获、数据处理和策略调整这一核心闭环,前者探测的是物理目标,后者捕捉的是公众意见和社会热点。

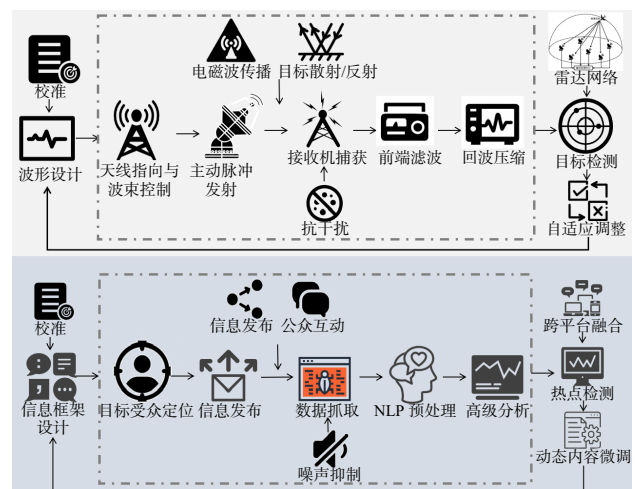


图 2 雷达与社会雷达主要环节对应

1.2.3 社会雷达的优势和局限性

自“阿拉伯之春”以来,所谓“社会雷达”已在智慧城市与网络安全等应用中展现出快速感知与相应的能力:它能够在社交媒体、在线新闻和舆论场域中高效捕捉社会信号,实时识别突发事件与公众态势的阶段性变化,体现出显著的时效性与敏捷性。与受限于局部物理空间观测的传统传感器不同,社会雷达可在跨地域的虚拟空间中实现大范围感知,具备广覆盖与跨域联动的优势;更为关键的是,它不仅回答“哪里出

现问题”，还能揭示“公众在关心什么、态度如何演化”，从而为理解合规要求、社会情绪与潜在风险提供语义层面的支撑。然而，该范式亦存在固有局限：其数据源噪声大、虚假与夸张信息频发，易诱发“信谣传谣”；网络热度与客观风险并非同构，高热度议题未必对应高现实危害，致使资源调度与行动优先级可能失衡；更具挑战性的是，如何将“热搜话题”“情绪曲线”“求救信号”等抽象社会信号可靠地映射为无人系统可执行的约束与目标，仍属跨学科的难题。换言之，仅依赖社会雷达，如“闻其声却不见其路”：虽能捕捉情绪之洪流，却难以据此稳健评估地形、障碍与行动可行性。

具身感知为无人系统提供了与物理世界交互的坚实基础。通过整合视觉、雷达、触觉和惯性测量等多源传感，系统能够高精度刻画环境几何、动力学状态与交互细节^[24]，并在自动驾驶、服务机器人和无人机导航中取得了重要进展。但此类感知主要停留于物理层面，难以有效理解与利用动态法规、公众情绪与群体行为等社会语境信息，因而在任务优先级、合规性保证与社会接受度提升方面面临瓶颈。社会雷达的提出正是为弥补这一物理与社会认知缺口：其可基于社交媒体、新闻与群体反馈快速捕捉社会信号与趋势，从宏观层面提供对公众态势与潜在风险的早期预警，在智慧城市治理、网络安全与智能交通中展现出独特价值。然而，社会雷达自身存在数据真实性不足与难以直接落地为控制约束等问题，因此难以单独支撑高可靠落地。

由此可见，具身感知与社会雷达构成“物理可见”与“社会可预”的互补范式：前者确保无人系统“看得清”，后者使其“听得懂”。两者结合，方能支撑从场景理解到决策执行的端到端闭环。例如，城市无人车除需识别交通灯与道路几何外，还应实时感知临时管制与人群潮汐；又如，救援无人机在采集灾情影像的同时，应结合社交平台求助线索动态调整优先级与航迹规划。但融合并非简单相加，现实中至少面临3类核心挑战：其一，跨模态语义对齐，如何将“禁飞区通告”或“人群灾后求助”等社会语义要素，形式化为规划与控制中的约束条件或代价函数，仍缺乏统一而可验证的建模框架；其二，多源数据的时效性与可信度评估，物理传感往往提供即时、可校验的实证数据，而社会

雷达信息可能失真甚至恶意操纵，亟需建立面上任务的置信度与溯源机制；其三，物理代价与社会优先级的权衡，群体协同任务中需在路径最优与能耗最小等物理目标与人群安全、公平性与合规性等社会目标之间实现可解释、可审计的多目标优化。

1.3 Agentic AI: 主体性与群体自治

1.3.1 Agentic AI 的内涵与机制

代理式人工智能(Agentic AI)的核心不再是孤立任务的执行，而是围绕从目标到行动的端到端闭环。其通过目标生成、任务分解、工具调用、记忆与反思，以及编排与治理5大机制的协同，能够在复杂、多步骤且需多方协作的场景中展现稳健的自治能力。与仅调用工具以完成单一任务的传统AI Agent相比，Agentic AI更近似由多类“专职代理”构成的群体化系统，具备持久化记忆与动态编排能力，能够在不确定与开放环境下自适应地调整策略。

这一范式转变可追溯至思想史脉络。亚里士多德曾区分“技艺理性”与“实践理性”：前者强调如何把事做成，后者强调在特定情境下做正确的事。传统AI更贴近前者，而Agentic AI正迈向后者，它不仅完成任务，还能在环境约束和社会规则下做对的事。正因如此，Agentic AI被视为未来自主系统通向可信、可解释和可落地应用的关键路径。为进一步界定其本质，对比了生成式AI(generative AI)、AI代理(AI agent)、生成式代理(generative agent)与代理型AI(agent AI)的关键特征(表1)。生成式AI以内容生成为主，依赖大语言模型(LLM)实现一次性响应，通常缺乏对外部数据与工具的原生访问；AI代理在此基础上引入工具使用能力，完成面向特定目标的任务执行；生成式代理进一步整合记忆与规划，使其在相对封闭的环境中具备更高的稳健性与一定程度的自主性；而代理型AI则以目标驱动的多智能体协同为核心，通过可观察、可记忆、可反思与可编排的闭环，支持复杂工作流自动化与外部环境的持续交互^[25]。

1.3.2 Agentic AI 的发展和应用

近年来，Agentic AI正逐步展现出作为多角色自治与编排系统的独特优势，能够在复杂、跨域与多步骤任务中实现高水平的自主性与可控性。围绕其核心机制的研究与应用，正在由理论探索快速迈向多领域的规模化落地。

表 1 生成式 AI、AI 代理、生成式代理与代理型 AI 的特征对比

名称	核心功能	机制	结构	外部数据访问	关键特征
生成式AI	内容生成	提示→大模型→输出	单一模型	无 (除非额外添加)	反应性
AI代理	使用工具执行特定任务	提示→工具调用→大模型→输出	大模型+工具	通过外部API	工具使用
生成式代理	模拟类人行为	提示→大模型+记忆/规划→输出	大模型+记忆+行为模型	通常为封闭世界 (模拟输入)	可信性/ 自主性
代理型AI	复杂工作流自动化	目标→代理编排→输出	多智能体系统	协调的多智能体访问	协作能力

在科研领域, Agentic AI 与无人机系统的深度融合催生了“Agentic UAV”新方向^[26]。Jiang 等^[27]提出的 ARMAIT 框架将大语言模型与注意力机制结合, 实现多无人机在低空经济网络中的轨迹优化, 在实验中显著提升了带宽利用率和任务完成效率。与此同时, Agentic AI 在自主空天、任务分配与跨平台协同中的应用正快速推进, 智能体系统 CoordField^[28]由大型语言模型负责解释高层人类指令并转译为群体无人机可执行命令; 其提出的协调场机制用于引导编队运动与任务选择, 实现紧急任务的分散化和自适应分配。在产业实践中, Agentic AI 已进入大规模试点阶段。2025 年 6 月, 亚马逊宣布组建 Agentic AI 团队, 以提升仓储机器人在调度、分拣与协作中的智能化水平^[29]。同年 9 月, McKinsey 的行业洞察报告表明, Agentic AI 正成为先进制造、物流和工业系统的关键引擎, 其价值在于提供可解释、可验证的自治决策链路^[30]。此外, 制造企业开始在生产线上嵌入 Agentic AI, 通过工具链调用与自我反思机制提升工艺优化与容错能力^[31]。总体而言, Agentic AI 的应用呈现 3 大方向: 其一, 无人系统智能化, 包括无人机群协同、低空经济调度与自主对抗规划; 其二, 工业与物流优化, 涵盖仓储调度、生产线优化与供应链动态管理; 其三, 跨域决策支持, 通过记忆、反思与多源验证机制推动系统从“能做”走向“做得对”。

上述进展亦为具身感知与社会雷达融合中的关键难题提供了新路径。Agentic AI 能够将社会雷达捕捉的网络信息、规则与风险信号, 转化为可计算的目标与可执行约束, 并与具身感知产生的物理实证数据进行对齐与联动, 从而在合规性、安全性与效率之间实现动态平衡。不同于传统线性的感知、决策链条, Agentic AI 构建了覆盖目标生成、任务分解、工具链

调用以及记忆与反思的可审计工作流, 使社会雷达不再停留在事件与热点的检测层面, 具身感知亦突破几何与物理量描述的局限; 二者在同一框架下形成互补耦合, 支撑面向多主体任务的协同决策, 并持续提升结果的可解释性与可验证性。

据此, 提出具身社会感知框架, 以 Agentic AI 为中枢, 将物理感知与社会感知深度耦合, 构建面向无人系统的感知、认知、决策闭环。该框架既满足城市治理中的实时合规需求, 又能支撑应急救援的快速响应与资源优化调度, 并在群体协同与复杂联演任务体现更高的稳定性与协调性。

2 具身社会感知智能的基本框架

随着无人系统逐步深入人类社会, 单纯依赖物理层面的感知与控制已难以支撑复杂多变的人机、群体交互。在开放而动态的环境中实现自然、可信且合规的社会行为, 要求无人系统进一步具备对人类情绪状态、意图表达与社会关系结构的建模与理解能力。为此, 在明确具身感知、社会雷达与 Agentic AI 这 3 者内在关系的基础上, 提出具身社会感知智能框架(图 3)。其中, 具身感知作为系统的感知层, 侧重通过多模态传感与融合实现对物理环境与社会环境的统一表征; 社会雷达作为推理层的关键支撑, 强调对情绪与社会关系的理解、预测与不确定性评估; 而 Agentic AI 位于元控制层, 作为顶层认知中枢, 负责目标生成、任务分解与跨层调度。3 者协同构建由元控制层、感知层、推理层、执行层与反馈层组成的动态闭环, 旨在赋予无人系统在复杂人机共融环境中的高级自主决策与社会适应能力。

具身社会感知智能框架由 5 个层级构成: 元控制

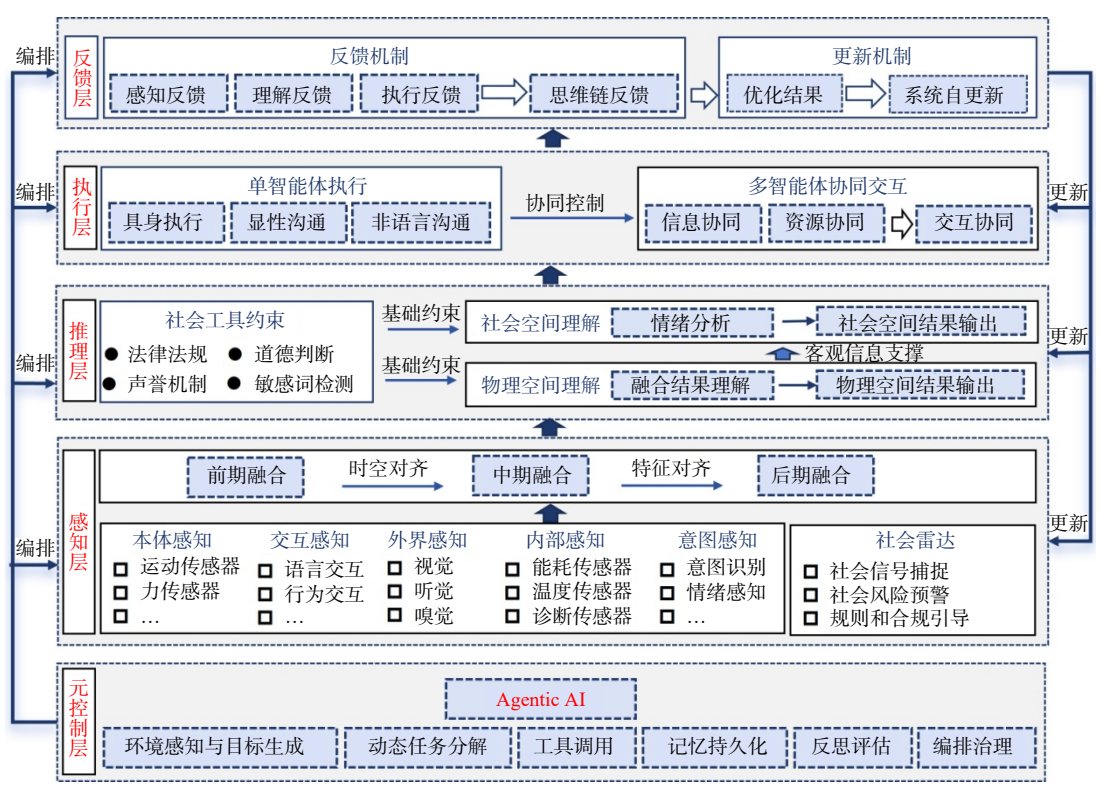


图3 具身社会感知智能框架

层位于顶层,以 Agentic AI 为核心,负责目标生成、任务分解与跨层调度,实现对系统的整体治理与编排;感知层作为系统与外界接口,整合本体、交互、外界、内部与社会 5 类感知信息,经由多模态融合形成一致且可更新的环境表征;推理层在统一表征之上开展情境化理解,一方面解析物理环境的几何与动力学特征,另一方面结合情绪与规范进行社会空间推理,并借助链式思维与检索增强生成等方法实现物理与社会的语义对齐与综合判读;执行层将抽象决策映射为可操作指令,既覆盖智能体的运动规划与交互控制,也支撑多智能体的协同分工与任务编排;反馈层对各层输出与环境响应进行采集与评估,提炼优化信号并持续更新模型参数,闭合自我进化的学习回路。5 个层级自上而下联动,使无人系统在复杂环境中既能“看得见、听得懂”,又能“做得对、持续学”。

3 具身社会感知系统的基本流程

在全球智能化浪潮的推动下,自主无人系统正从执行预设任务的工具,迈向能够在复杂、动态社会环境中展现自主适应与合规行为的智能体。为达成这

一目标,系统既需具备对物理世界的高精度感知与建模能力,也需理解和遵循社会规范。本文提出的“具身社会感知智能”框架,以元控制层为统筹中枢、联动感知层、推理层、执行层与反馈层,形成动态工作流,赋予无人系统从多模态信息获取到合规性行为输出,再到经验反馈与自我进化的完整能力。其核心逻辑在于引入基于 Agentic AI 的元控制层,将社会雷达与具身感知整合进统一决策闭环,使无人系统逐步成为具备物理实证与社会适应能力的“具身社会智能体”。

该框架的整体工作流程由元控制层的顶层设计与驱动启动。其元控制层采用 Agentic AI 技术,将单一系统或系统集群塑造为具备自治性的智能体:首先接收由人类下达的高层宏观目标;随后利用规划与分解算法,将复杂总体目标拆解为逻辑清晰、可执行的子任务;继而完成工具链选择与资源编排,决定后续流程中启用的算法模型与外部数据接口;并对感知、推理、执行与反馈 4 层进行持续治理与协调,以保证全链路的一致性与效率。该过程体现出自顶向下的策略制定与自底向上的闭环校正相结合的机制,使系统在任务推进中实现动态监督、协同与优化,确保目标可达与路径可审计。

感知层作为闭环的起点,接收来自系统内外海量且异构的原始信息流,涵盖5大感知维度:其一,本体感知通过运动与力觉传感器获取自身姿态与受力状态;其二,交互感知依托语言与行为交互模块解析人类的直接指令与隐性意图;其三,外界感知利用视觉、听觉与嗅觉等多模态传感捕捉物理环境的动态变化;其四,内部感知借助能耗、温度与诊断信号监测系统健康;其五,意图感知面向意图识别与情绪感知,结合语音语调、文本语义、面部表情、姿态/手势以及必要的生理信号等证据,对人类目标、偏好与情绪状态进行联合建模与估计。至关重要的社会感知通过网络信息获取社交媒体情绪状态。面对庞杂多模态数据,感知层采用以特征空间特征对齐为核心的多模态融合范式,并非简单叠加而是分阶段处理:首先在数据源头进行前融合以处理天然耦合信号,其后通过时空对齐将不同频率与时间戳的数据映射到统一时空基准,继而在中融合阶段从各模态抽取深层表征并以对齐算法将视觉(图像)、语音(声学)与文本(语义)等映射至统一的高维共享隐空间,最终在后融合环节对不同分支的决策级或语义级输出进行加权组合。由此产生的结构化、语义丰富的“感知融合结果”,既提供对物理世界的精确刻画,也包含对社会情境的初步理解,为后续推理层的情景化推断与约束构建提供高质量、高信息密度的决策依据。

接收感知层的融合表征后,推理层启动其核心的认知与决策过程,旨在基于当前信息生成满足物理规律与社会规范的行动意图。该层以思维链与检索增强生成为基础架构,由社会工具约束、物理空间理解和社会空间理解3模块协同作用:其一,物理空间理解对融合表征中的客观要素进行解析,包括障碍物识别、行人轨迹预测与交通标志的语义-物理映射,形成关于环境状态的可检验结论;其二,社会空间理解聚焦人类与群体要素,依据意图与情绪线索、社交距离与互动模式、潜在群体行为趋势等信息,输出关于社会氛围与人际关系的情景化判断;其三,社会工具约束充当内置的法理-伦理顾问,通过检索增强生成从外部知识库实时检索并整合法律法规、行业规范、组织合规策略与敏感规则,对前2者的结论进行一致性校验、冲突消解与边界设定,从而在策略层面实施硬约束与软指导。3者并行迭代、相互制约:物理空间

理解为社会空间理解提供可观测事实基础,社会空间理解为物理可行解提供人与群体层的可接受性评估,而社会工具约束则据以筛除违法或不当方案并重加权可行集(例如,当物理模块给出捷径方案而约束模块判定其涉及私域侵入或交通违规时,该方案即被否决)。思维链技术在此过程中扮演了关键角色,它使推理过程不再是一个黑箱,而是能够生成一步步清晰的推理路径。在可解释推理框架的支撑下,推理层形成逐级可审计的推理流程与可复核的知识溯源,最终给出物理-社会空间综合理解结果,即兼具事实依据、逻辑一致、法理合规与伦理可接受性的高阶决策意图,为下游执行层提供可操作、可验证且具有边界意识的行动指南。

推理层输出的决策意图被输入至执行层,用以指导具体行为执行。执行层的核心任务是将这类高阶、抽象的意图经由动作解码器与控制栈转译为可操作的物理动作和与人机、机机交互行为,其流程包含单智能体执行与多智能体协同2个层面。对于单智能体执行,系统产生细粒度控制指令:一方面进行具身执行,通过运动规划、在线路径重规划与避障策略驱动机构运动;另一方面实施隐性沟通以向周围的人类或其他智能体表示即将动作的意图;在必要情景下通过显性沟通(语音提示、解释与警告)实现直接人机交流。当系统由多个无人单元构成时,元控制层通过治理与编排将若干独立的单智能体执行上升为多智能体协同。这种协同并非简单的指令广播,而是通过3个层面的紧密配合实现:信息协同确保各智能体共享关键的感知数据,形成统一的环境态势图;资源协同根据各单元的能力、位置和当前负载,动态分配任务;交互协同则通过预设的协议或实时协商,确保多个智能体在物理空间中的行动能够互不干扰、高效配合,共同完成复杂任务。因此,执行层的输出是一系列精确的执行信号,直接驱动系统的电机、扬声器、显示屏等硬件单元,将智能决策最终外化为在物理与社会世界中的实际影响。

为了实现系统的持续学习与进化,反馈层闭合了整个工作流程的环路。其输入涵盖感知、推理与执行3个前序层面的全过程数据与任务结局,既包括感知准确性、推理合理性与执行达成度等全局指标。反馈层内部由反馈机制与更新机制构成。反馈机制负

责对多源异构信息进行采集、对齐与融合,既关注任务的整体成败,也追踪各环节的过程性细节:感知层可能出现的误检、漏检与配准偏差,推理层可能暴露的逻辑矛盾与规则冲突,执行层记录的实际动作与期望指令之间的轨迹或时序偏差等。上述信息被统一汇聚,形成具备上下文关联与可追溯性的综合反馈。随后,更新机制启动,采用结构化摘要与因果归纳等方法,将冗长分散的反馈提炼为简洁、可操作的优化结论,明确“何处需要改进、为何产生偏差、应采用何种策略修正”。该结论作为自适应更新的依据,分别定向用于感知层、推理层与执行层。通过反馈、提炼、更新的闭环迭代,无人系统得以在每次任务后持续吸收经验、修正误差并适应环境漂移,使其具身社会感知智能实现性能提升,进而在复杂真实场景中表现出更高的效率、稳健性与社会协同度。

4 结论

围绕具身社会感知智能这一新概念,系统梳理了具身感知、社会雷达与 Agentic AI 的发展脉络与现实需求,指出单一物理感知或单一社会感知在复杂开放环境中均难以支撑自主无人系统的可靠运行。在此基础上,构建了以 Agentic AI 为中枢、涵盖元控制、感知、推理、执行与反馈 5 个层级的具身社会感知智能框架,并给出了从信息采集、多模态融合到决策生成与自适应更新的完整流程。该框架面向城市治理、应急救援和社会安全等典型场景,强调物理“看得见”和社会“听得懂”的协同,通过合规性建模、社会风险预警与群体协同优化,提升自主无人系统的任务适应性与工程可落地性。

需要指出的是,当前工作仍存在若干有待深入的问题。在复杂开放环境中,社会信号的建模与噪声抑制仍较为初步,具身感知与社会雷达的融合机制在跨场景迁移、长期稳定性和极端情境下的鲁棒性方面尚需系统验证;合规性建模目前更多依赖规则与专家知识,如何在保证安全前提下引入自适应学习与在线更新机制,也是未来需要重点解决的难题。此外,Agentic AI 的多智能体协同在安全边界、责任划分与人机协同治理方面,仍需进一步引入法律、伦理与政策层面的系统研究。

总体而言,本文为自主智能无人系统从“任务执行体”迈向“社会智能体”提供了一个具有理论完整性与应用前景的方法论起点,为后续在规则建模、可信评估与跨学科协同方面的深入研究奠定了基础。

参考文献(References)

- [1] 房超. 从俄乌冲突看智能化战争的典型特征与制胜规律[J]. 国防科技工业, 2025(3): 48–54.
- [2] 白孟宸. 人工智能时代, 战争形态是否会彻底改变?——俄乌冲突中的认知战(三)[EB/OL]. (2025–02–26) [2025–09–04]. https://www.guancha.cn/baimengchen/2025_02_26_766334.shtml?
- [3] 蓝德智库. Palantir 大数据技术在乌克兰战场的应用[EB/OL]. (2023–09–15) [2025–09–04]. https://www.sohu.com/a/720747594_121294014?
- [4] Greenwood F, Nelson E L, Greenough P G. Flying into the hurricane: A case study of UAV use in damage assessment during the 2017 hurricanes in Texas and Florida[J]. PLoS One, 2020, 15(2): e0227808.
- [5] Imran M, Alam F, Qazi U, et al. Rapid damage assessment using social media images by combining human and machine intelligence[J/OL]. (2020–04–14) [2025–09–04]. <https://arxiv.org/abs/2004.06675>.
- [6] Liu C, Sziranyi T. Active wildfires detection and dynamic escape routes planning for humans through information fusion between drones and satellites[C]//Proceedings of IEEE 26th International Conference on Intelligent Transportation Systems (ITSC). New York: IEEE, 2023: 1977–1982.
- [7] von Buren H F, Cardoso M M. Civilian drones as an exogenous player on the search and rescue teams process. case study analysis of the Rio grande do Sul state flood using FRAM[C]//Proceedings of 35th European Safety and Reliability Conference (ESREL 2025) and the 33rd Society for Risk Analysis Europe Conference (SRA-E 2025). Singapore: Research Publishing Services, 2025: 1–6.
- [8] Li S H, Xu J L. Multiaxis force/torque sensor technologies: Design principles and robotic force control applications: A review[J]. IEEE Sensors Journal, 2025, 25(3): 4055–4069.
- [9] Tang J, Yuan X, Li S. Visual–tactile fusion and SAC-based learning for robot peg-in-hole assembly in uncertain environments[J]. Machines, 2025, 13(7): 605–619.
- [10] Wu S, Zheng Y, Ye K, et al. Sound source localization for unmanned aerial vehicles in low signal-to-noise ratio environments[J]. Remote Sensing, 2024, 16(11): 1847–1873.
- [11] Li J, Zhang X, Li J, et al. Building and optimization of 3D semantic map based on LiDAR and camera fusion[J]. Neuro-

- computing, 2020, 409: 394–407.
- [12] Mohammadi M A. Energy consumption and temperature sensing for health-aware operation of embodied mechatronic systems[J]. IEEE Sensors Journal, 2024, 24(12): 15210–15222.
 - [13] Gao Y, Shen X, Yang Z, et al. Structural health monitoring sensors for bearings and gears in embodied mechatronic systems[J]. IEEE Sensors Journal, 2024, 24(15): 18320–18335.
 - [14] Zhuang Y, Ding H, Arnold P, et al. Flexible facial expression sensors and high-density EEG/EMG arrays for intention decoding in social robots[J]. IEEE Transactions on Human-Machine Systems, 2024, 54(5): 855–868.
 - [15] Behzad A, Di Stefano G, Miura K, et al. Non-contact crowd flow and social behavior sensing *via* WiFi, millimeter-wave radar, and infrared arrays[J]. IEEE Transactions on Intelligent Robots, 2024, 40(7): 2153–2168.
 - [16] 联合国人权理事会. 自主武器系统[R]. 日内瓦: 联合国人权理事会, 2024.
 - [17] Schramm W L, Porter W E. Men, women, messages, and media: Understanding human communication[M]. 2nd ed. New York: Harper & Row, 1982: 45–52.
 - [18] Wang F Y. Social media and the jasmine revolution[J]. IEEE Intelligent Systems, 2011, 26(2): 2–4.
 - [19] Mathieu J, Fulk M, Lorber M, et al. Social radar workflows, dashboards, and environments[C]//NATO HFM-201 Specialist Meeting on Social Media: Risks and Opportunities in Military Applications. Paris: NATO, 2012.
 - [20] Schramm W L. Men, messages, and media: A look at human communication[M]. New York: Harper & Row, 1982: 30–36.
 - [21] Mathieu R, Maybury M, Laboreiro G. Social radar: Leveraging social media for predictive analytics[J]. IEEE Intelligent Systems, 2012, 27(6): 22–28.
 - [22] Fan F, Liu X, Wang F-Y. Social radars for social vision of intelligent vehicles: A new direction for vehicle research and development[J]. IEEE/CAA Journal of Automatica Sinica, 2024, 11(5): 1050–1063.
 - [23] Cao Z, Li M, Wang X. Traffic sentiment analysis for intelligent transportation systems: A social radar approach[J]. Transportation Research Part C, 2014, 44: 240–252.
 - [24] Yang L, Chen X, Zhang Z W, et al. A cross-modal alignment method based on adaptive feature aggregation and spatial fusion mechanism[C]//Proceedings of 4th International Conference on Electronic Information Engineering and Computer (EIECT). Shenzhen: IEEE, 2024: 1–4.
 - [25] Sapkota R, Roumeliotis K I, Karkee M. AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges[J]. Information Fusion, 2025, 126: 103599.
 - [26] Sapkota R, Roumeliotis K I, Karkee M. UAVs meet Agentic AI: A multidomain survey of autonomous aerial intelligence and agentic UAVs[J/OL]. [2025-06-13]. <https://arxiv.org/abs/2506.08045>.
 - [27] Jiang F, Dong L, Pan X, et al. Agentic AI empowered multi-UAV trajectory optimization in low-altitude economy networks[J/OL]. [2025-08-27]. <https://arxiv.org/abs/2508.16379>.
 - [28] Zhang T, Tian Y, Lin F, et al. CoordField: Coordination field for agentic UAV task allocation in low-altitude urban scenarios[EB/OL]. [2025-05-01]. <https://www.arxiv.org/abs/2505.00091>.
 - [29] Investopedia Staff. Amazon launches agentic AI group to enhance its warehouse robots[EB/OL]. (2025-06-05) [2025-10-07]. <https://www.investopedia.com/amazon-launches-agentic-ai-group-to-enhance-its-warehouse-robots-reports-say-11749004>.
 - [30] Jautelat S, Niemann J, Barg J, et al. Empowering advanced industries with Agentic AI[EB/OL]. (2025-09-08) [2025-10-07]. <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/empowering-advanced-industries-with-agentic-ai>.
 - [31] Lee J, Su H Q. Agentic AI for smart manufacturing[J]. Manufacturing Letters, 2025, 46: 92–96.

Embodied social perception intelligence for unmanned systems: Workflow and framework

FAN Lili¹, SHI Guangyu², CHEN Kaige¹, LI Chuang³, ZENG Changxian⁴, DENG Fang^{1*}

1. School of Artificial Intelligence, Beijing Institute of Technology, Beijing 100081, China

2. School of Artificial Intelligence, Beijing Institute of Technology, Zhuhai 519088, China

3. Military Intelligence Institute of the Academy of Military Sciences of the Chinese People's Liberation Army, Beijing 100091, China

4. School of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China

Abstract Autonomous intelligent unmanned systems operating in real-world open environments—characterized by dynamic complexity, multi-agent coupling, incomplete information, and strong social constraints—face critical challenges such as insufficient compliance modeling, limited social risk perception, complex collaborative conflicts, and delayed abnormal response. To address these issues, this paper proposes an Embodied Social Perception Intelligence Framework, which integrates embodied perception (including proprioceptive, internal, exteroceptive, interactive, and intention perception) with social radar, and introduces Agentic AI as a top-level decision-making and control mechanism to achieve multi-level and autonomous cognitive decision-making. The framework adopts a five-layer architecture—perception, reasoning, execution, feedback, and meta-control—establishing a dynamic closed loop from multimodal perception to compliant behavior generation. By fusing physical and social environmental information, the proposed framework significantly enhances the task adaptability, collective coordination efficiency, and compliance reliability of autonomous intelligent unmanned systems in complex and uncertain scenarios such as urban governance, emergency rescue, and social security. This work provides a new technical pathway toward trustworthy, explainable, and sustainable autonomous intelligent systems.

Keywords embodied perception; social radar; agentic AI; autonomous intelligent unmanned system ●



(责任编辑 傅雪)