



中国人工智能学会-联想 蓝天科研基金项目申报主题

蓝图汇众智 天高任鸟飞

中国人工智能学会-联想 蓝天科研基金

由中国人工智能学会与联想共同设立，面向中国高校和科研院所青年学者的科研计划

发布背景

- 响应政府号召，推动产学研协同发展。
- 联想集团持续支持高校学者的科研探索，共同解决前瞻技术挑战。
- AI是联想关键战略方向， CAAI是我国智能科学技术领域唯一的国家级学会。

课题方向

- 大模型与智能体 (8个)
- 具身智能 (6个)



前沿探索类：10万/1年，鼓励前瞻创新，探索前沿技术机会，成果以学术交流及研究报告为主。



单点突破类：30万/1年，创新地解决行业技术挑战，共同推动成果应用。

专项一：大模型与智能体

1	前沿探索类	AI Agent硬件载体交互方式和形态探索和产品创新设计
2	前沿探索类	基于心理根逻辑的智能体交互体验示范研究
3	单点突破类	AI Agent交互研究
4	单点突破类	基于模型表示的端云多模态大模型路由技术研究
5	单点突破类	基于大语言模型的全球供应链网络产能规划与决策智能体研究
6	单点突破类	基于 Human-in-the-loop 的设备端交互智能体用户体验提升研究
7	单点突破类	基于认知心理的 Agent OS 隐喻与交互范式研究及创新体验设计
8	单点突破类	多智能体系统运行状态追踪与协同响应机制研究

课题类型：前沿探索类

1、研究背景：

AI Agent软件产品技术日趋完善和成熟，智能程度快速提升。目前市场上尚无成功商业化的专门针对AI Agent而设计的硬件产品，主要是缺乏杀手级的使用场景和应用案例，且AI agent现有概念产品在算力使用顺畅度、落地时限性方面存在诸多问题，无法实现商业化。

2、研究内容：

以硬件为载体，基于AI Agent的功能、交互和应用等方面的时代特点，推导出新的产品形态、功能和交互，应用在以人为核心的场景中，如办公、陪伴等等，以此探索新产品品类的可能性。

1. 用户与场景洞察

- 研究AI Agent时代下用户与设备交互的行为变迁（办公、出行、家庭陪伴等）
- 提炼“软硬融合”交互中的未满足需求与未来趋势洞察。

2. 硬件形态与交互原型探索

- 研究AI Agent载体的核心特征（可感知、可响应、具备情境理解）
- 探索不同硬件形态（可穿戴/伴随设备/桌面终端）的交互逻辑
- 形成3~5个未来硬件+软件交互概念模型

3. 设计方向与验证

- 产出面向未来的AI Agent硬件+软件交互趋势报告
- 选取1~2个典型场景进行交互原型或视频化验证

3、研究目标：

针对随身携带的个人数字分身设备新形态提出的一种全新软硬件结合的交互方式、产品形态洞察并定义典型应用场景提出具有成为杀手级的创新点，并通过用户体验验证。

- 提炼AI Agent时代下硬件产品创新趋势与交互模式演进路径；
- 提出可指导未来AI设备设计方向的交互设计原则与形态演化模型；
- 输出具备purchase driver潜力的创新概念与设计验证成果。

4、产出及交付物：

交付类型	内容说明	可验收形式
研究报告	《AI Agent硬件形态与交互趋势探索报告》	完整PDF报告（≥30页）
设计成果	3~5个AI Agent硬件+软件交互概念方向（附形态草图或渲染）	概念设计合集文档
验证成果	典型场景交互演示（视频或交互说明文档）	视频演示+交互说明PPT

项目名称：基于心理根逻辑的智能体交互体验范式研究



课题类型：前沿探索类

1、研究背景：

当前智能体在情感交互与心理理解层面仍存在明显局限，如难以捕捉人类潜意识动机与复杂心理防御机制。基于心理根逻辑的智能体交互研究，可填补现有技术对深层心理机制建模的空白。

2、研究内容：

AI交互如何在新模态下提供各种心理体验与支持，包括但不限于：新型的沟通范式，认知效用，情感维护，性格支援，信任与依赖，行动助推，文化传导，习惯搭建，思维方式重塑，品格修养，社会融入等；要系统性考虑输入（用户信息）和输出（个性化应答）的精准适配，提升AI交互的粘性和频次。

3、研究目标：

形成一套基于用户心理特征的AI交互策略；实现硅基智能体（L3 自主执行/L4 自主创造）的情感真实性突破。在企业提供的场景中做验证测试可提高客户粘性。

4、产出及交付物：

- 全球学术界，业界关于智能体交互体验的设计趋势分析
- 典型场景的应用范例和相应的效果测试报告
- 通用的基于用户心理特性的AI交互体验设计指导

课题类型：单点突破类

1、研究背景：

当前，大模型技术持续突破，AI Agent 能力快速迭代，智能交互正从“指令响应”向“主动协同”演进。L4 级别 AI Agent 凭借高度自主决策与跨场景任务闭环能力，打破传统交互边界，推动人机协作模式从用户单向诉求，转向 Agent 主动理解、预判并协同完成复杂目标，同时还具备挖掘用户任务背后更深层需求并自主提出推荐方案的能力。这些对智能体的交互范式创新提出迫切需求。在领域智能体（如面向销售、企业内部办公等场景的平台）面临用户类型多、信息繁杂、流程割裂等问题。用户常需在多个子系统模块间切换，手动整合数据，传统被动交互模式已难以适配复杂销售场景。基于此，对这些内部产品进行智能化改造，打造 AI 原生产品，期望通过AI Agent L4的高度自主决策和跨场景任务执行能力来解决PC端多用户、多场景复杂工作流，既是顺应技术演进的必然选择，也是破解业务痛点的关键路径。

2、研究内容：

- 高度自主与用户干预的边界：既让 Agent 充分发挥主动决策优势（如自动处理重复性子任务），又保障用户对关键环节的控制权。
- 多模态交互形式的自然融合：语音、手势、环境感知等多模态输入形式对触发条件、信息密度、用户心智都有差异。如何让多模态交互自然衔接（从语音唤醒到手势调整），根据任务类型、场景环境如何动态选择模态组合，确保产品的高可用性。
- 复杂任务闭环的反馈：多步骤、跨系统任务时如何向用户传达步骤透明、决策可解释以及异常状态（如权限问题）处理的预案

3、研究目标：

基于L4+知识检索能力的交互范式研究，给出交互原型和场景洞见，为公司智能体交互设计提供创新机会点，供项目具体设计落地提供具参考价值的设计方案

4、产出及交付物

交付类型	内容说明	可验收形式
研究报告	AI Agent交互趋势分析报告	完整pdf报告（≥20页）
设计成果	3~5个场景 完整的交互设计方案（figma设计），并标注所有场景背景信息、交互细节、设计亮点等说明。	Figma 设计稿
验证成果	3~5个场景完整的高保真demo演示视频。场景不要重复，需要展示多模态交互形式、多用户复杂工作流、多个子agent 协作演示。	至少3个视频演示+说明
场景	1、距离会议开始还有 5 分钟，当你起身去拿纸质资料时，AIOS 通过摄像头捕捉到动作，结合日历中“产品周会”的日程，自动启动会议助手开始对成员分发资料，会议中。通过多模态感知，实时捕捉每一个参会人员输入，例如声纹识别自动标注其身份，智能体同步调取他负责模块的技术文档，辅助他快速同步发言内容。会议后自动沉淀多维度会议纪要，并将todo同步至每一个负责人的日历中。 2、本月14~15日我需要拜访武汉某企业客户请帮我梳理近两年详细客户资料，并预定北京武汉的往返机票。 3、整理本周营销主题会议结论，生成关键营销建议，创建图文营销方案。	无 6

课题类型：单点突破类

1. 研究背景：

跨模随着大语言模型（LLM）、大多模态模型（LMM）的爆发式发展，模型的能力边界持续拓展——从单一文本理解到跨模态交互（文本、图像、语音等），从通用场景适配到垂直领域深耕，从云端大规模部署到端侧轻量化落地，形成了多维度、多层次的模型生态。然而，这种“模型爆炸”也带来了新的挑战：不同模型在任务适配性、部署成本、响应时延等方面呈现显著差异，例如，云端大模型虽具备强大的复杂任务处理能力，但存在高算力消耗、网络依赖强的问题；端侧轻量化模型虽能实现低时延、低成本部署，却在复杂推理、多模态融合等场景中能力受限。

尤其在多模态场景中，输入信息的异构性（文本描述+图像数据、语音指令+视频流等）进一步加剧了路由决策的复杂性：不仅需要解析不同模态的语义内涵，还要综合评估任务整体复杂度与模型的跨模态处理能力。此外，随着端云协同架构的普及，如何在本地资源约束与云端能力优势之间找到最优调度策略，成为平衡效率、成本与体验的核心命题。大模型智能路由技术正是在这样的产业需求与技术演进中，逐渐成为连接多样化模型与场景化需求的关键枢纽。

2. 研究内容：

与基于文本的大模型路由的核心差异，多模态大模型路由的输入包含文本和图像两种模态，而基于文本的大模型路由仅以文本为输入，因此在模态处理逻辑上存在本质区别。主要技术难点包括：

- 跨模态内容的综合理解与复杂度评估

文本与图像的语义表征方式差异显著（如文本依赖上下文逻辑，图像依赖视觉特征），需实现对两种模态内容的语义融合理解，并综合判断整体问题的复杂度，这对路由决策的准确性提出更高要求。

- 图像内容的低算力高效分析

模型路由作为模型调用的前置环节，需严格控制自身的算力消耗及时延（避免增加整体调用开销）。但图像内容的分析（如图像特征提取、语义识别）往往算力需求较高，如何在低算力约束下实现对图像内容的高效解析，是关键挑战。

3. 研究目标：

- 路由模型在现有公认的多模态榜单（包括MMBench、MMStar、MMMU、MMMU-Pro、MegaBench，以及数学MathVista、图表InfoVQA、视频VideoMMMU、GUI Agent AndroidWorld）的性能评测上相比SOTA平均提升5%以上
- 路由模型需要支持多个多模态模型间路由（线上+本地），同等/更优精度下，路由成本比基线平均节约 ≥ 10 个点，路由实时性指标为4060及以上显卡环境中，通用图文指令平均时延 ≤ 1 s。
- 支持自动进行新大模型模型能力边界探测、路由模型可在1天内完成自训练更新，保证新的大模型接入后整体模型路由性能为正向提升。

4. 产出及交付物：

- 发表CCF-A类或联想认可的期刊会议论文1-2篇。
- 完成多模态模型路由相关技术的报告2篇。
- 完成多模态模型路由的算法原型，交付一套可迭代、可演进的源代码和文档。

课题类型：单点突破类

1. 研究背景：

随着全球化生产网络日益复杂，企业在全球供应链产能规划中面临严峻挑战。传统的、依赖静态规则（如成本最低）的优化算法已难以应对多变的生产需求、地缘政治和关税壁垒等动态因素。为应对此挑战，我们提出构建一个基于大语言模型（LLM）的智能体（Agent），旨在实现对全球供应链产能规划的深层次、智能化分析与决策。初步研究发现，项目的核心技术瓶颈在于如何引导大模型超越其通用知识，深度理解我们特定的业务场景，以挖掘出超越传统算法的、更具洞察力的产能分配方案。

2. 研究内容：

本项目将围绕智能体的两大核心能力进行单点突破研究，并以第一点为最优先攻关方向：

- 融合领域知识的LLM智能规划研究：研究的核心在于解决如何利用LLM发掘超越传统优化模型（如仅基于成本和最大产能策略的“小模型”）的潜在最优产能分配方案。传统算法难以量化和应对如“美国关税频繁变动”这类地缘政治风险因素，而这正是LLM的优势所在。然而，研究中必须正视一个关键差异：大模型（如Deepseek、Qwen3）自身的通用分配能力与其在我们特定业务场景中所需达成的效果存在显著鸿沟。通用LLM基于其对整个供应链行业的宏观理解进行推荐，这种推荐往往是宽泛且不精确的。我们的场景则要求大模型必须深度理解自身的业务逻辑、工厂分布、产品成本结构和面临的特定风险，并在此基础上进行推理和规划。因此，本研究不寻求通用解决方案，而是聚焦于：如何将我们独有的业务知识、隐性规则和动态风险（如关税、地缘政治）有效注入并“锚定”LLM的推理过程，使其从一个“行业通才”转变为精通我们业务的“专家”，从而生成真正符合我们特定需求的、更优的、具有前瞻性的产能分配策略。
- 意图驱动的混合任务框架研究：解决智能体在多轮对话中对用户深层、复合意图（如：先查产能，再查成本，最后要求基于前两者进行分配）的精准识别问题。研究如何将该意图准确转化为对数据库查询（Text-to-SQL）和调用内部分析模型的混合任务流，为上述的智能规划提供准确、可靠的数据输入。

3. 研究目标：

- 智能规划能力：构建一个融合了本企业特定业务知识的LLM智能规划模块。在包含模拟地缘政治风险（如关税突变）的测试场景下，其生成的产能分配方案，在综合评估（风险规避、成本节约、供应链韧性）上需超越传统优化算法至少15%。系统必须能以报告形式，清晰解释其优化建议背后所考量的特定业务因素和风险逻辑。
- 意图驱动框架：在连续三轮以上的复杂对话场景下，实现对用户最终复合意图的识别准确率 $\geq 95\%$ ，并确保其分解生成的数据库查询SQL准确率 $\geq 98\%$ 。

4. 产出及交付物：

- 知识产权：相关技术专利交底书1-2篇，高质量学术论文1篇。
- 算法与框架：
 - 一套锚定企业业务知识的LLM智能规划模块及设计文档（核心产出）。
 - 一套意图驱动的混合任务框架源代码及接口文档。
- 系统原型：可演示上述研究目标的智能决策系统Demo一套。

课题类型: 单点突破类

1、研究背景:

在多模态大模型驱动下, 端到端 GUI 智能体成为设备自动化核心方向, 但存在三大瓶颈: 过度依赖离线完美演示数据, 缺乏在线纠错与动态学习能力, 个性化场景准确率不足; 强化学习面临 “探索 - 利用” 平衡难题, 用户指导与算法优化融合路径模糊, 难以转化人类经验为模型迭代动力; 经典 “截屏 - 分析 - 执行” 串行模式效率低下, 影响用户体验。为此, 将 Human-in-the-loop 机制深度融入强化学习闭环, 构建用户指导与算法优化的协同体系, 同时赋予智能体 “单次分析生成连续动作” 能力, 是突破瓶颈、提升自动化效果与用户体验的关键路径。

2、研究内容:

任务1: 创新构建 GUI-Agent 的 “熟路模式” 框架, 基于用户指导的历史交互数据构建个性化任务知识图谱, 实现从模糊指令到精准操作的动态适配, 使智能体能够快速复用人类经验应对相似场景, 显著提升个性化任务的处理准确率。

任务2: 研发 “单次场景识别 - 连续动作生成” 的端到端执行引擎, 通过多模态 GUI 特征的一次性全局解析与动作依赖关系建模, 并且形成基于记忆知识图谱指导智能体行动, 替代传统 “截屏 - 分析 - 执行” 的串行循环, 大幅减少交互延迟与系统资源消耗, 从根本上优化智能体的任务执行效率与用户体验流畅度。

3、研究目标:

构建用户引导系统及适配模型, 通过采集特定场景下用户任务操作数据实现设备端模型微调, 提升该场景任务成功率 (SR) ; 在 WAA 数据集上单独 Working Domain 上, 基于用户引导模型的优化方案实现 **86% 以上 SR**, 且在其余 WAA 案例及 OS-World 数据集无性能退化, 验证适配性与稳定性; 在相同环境下, 使方案执行效率较调用 GPT-o1 提升 **2 倍以上**, 达成性能与效率双重优化。

4、产出及交付物:

- 首个将Human-in-the-loop与序列化动作执行结合的GUI-Agent框架。
- 在HCI或AI顶会, 发表1-2篇论文, 获得国家授权专利1-2项。

课题类型：单点突破类

1. 研究背景：

随着 AI Agent 技术的发展，传统人机交互范式正经历深刻变革，不仅为用户带来全新体验，也对交互设计提出新挑战。当前 Agent OS 的设计实践中，缺乏基于用户认知的交互隐喻与操作范式，对于用户如何理解 Agent 等问题尚无系统化的理论或方法。同时，用户在面对全新界面时，常因认知负荷与学习门槛高而产生挫败感，进而影响系统的接受度与使用效率。因此，需探索更符合认知机制的设计方案。

2. 研究内容：

- 基于认知心理学，构建用户理解 Agent OS 的心智模型，定义交互隐喻与操作范式。通过设计和测试多种具有代表性的交互隐喻，分析其在不同任务场景中对用户理解度、接受度、操作效率及认知负荷的影响，最终形成一套以用户认知为核心的交互范式体系。
- 根据交互范式，针对文件管理和Memory管理等典型功能场景开展创新设计，优先考虑当前主流设备形态（PC、平板、手机等）。

3. 研究目标：

- **交互隐喻与范式有效性验证：**基于用户对界面的理解程度、操作负荷、可用性与信任感进行系统测量，达到以下量化指标或其他同等指标（理解准确率 $\geq 85\%$ ，意图表达准确率 $\geq 90\%$ ，NASA-TLX ≤ 50 ，SUS ≥ 80 ，信任感评分 $\geq 4.0 / 5.0$ ）
- **创新设计方案有效性验证：**针对文件管理与Memory管理两个场景的创新设计，通过用户行为与体验反馈综合验证其实际效果，达到以下量化指标或其他同等指标（文件检索效率提升 $\geq 40\%$ ，满意度评分 $\geq 4.5 / 5.0$ ）

4. 产出及交付物：

- **设计：**输出一套包含交互隐喻和范式定义的设计规范；输出典型功能（文件管理、AI memory）设计方案以及交互原型。
- **论文：**发表联想认可的 CCF-A 类会议、期刊论文 1-2 篇。
- **报告：**提交“符合用户认知的 Agent OS 交互隐喻和范式定义”、“Agent OS 界面设计指导原则”等报告 1-2 篇。
- **专利：**提交专利 2 项。
- **预计时间点：**其中 T+3 产出交互隐喻和范式，T+7 完成设计方案及交互原型。

课题类型：单点突破类

1、研究背景：

随着多智能体系统在复杂任务协同中的广泛应用，系统内部智能体间的动态交互与状态依赖日益增强，导致运行过程中个体异常引发连锁效应，影响整体任务的正确性与稳定性。当前多数系统缺乏对运行过程的细粒度追踪机制，难以实现异常行为的有效感知与传播路径分析。同时，在应对不确定性时，系统在自动决策、人工介入协同与自适应修复方面仍存在能力短板。因此，亟需构建一套具备状态可观测性、异常可解释性与处置灵活性的通用支撑体系，以提升多智能体系统的鲁棒性与可信性，支撑其在多样化应用场景中的可靠运行。

2、研究内容：

本项目聚焦多智能体系统的运行状态追踪与异常协同应对机制，研究内容包括：构建基于因果建模的异常根因定位框架，实现“异常感知—因果关联—责任归属”的全流程溯因，并引入反事实推理揭示深层因果机制；建立多维度归因能力评测体系与基准数据集，对溯因的定位准确性、因果正确性与责任一致性进行系统量化；构建基于溯因与评测结果的反馈优化机制，形成从根因识别到策略生成与验证的闭环流程，推动多智能体系统向高可靠性与自我进化方向发展。

3、研究目标：

构建一个通用、可扩展的多智能体运行支撑系统，实现对智能体协同过程的透明化追踪与异常响应能力。具体目标包括：实现端到端运行状态的毫秒级采集与重构，异常事件检测延迟低于1秒，异常归因能力在公开数据集Who&When上较SOTA方法提升至少5%，系统在获得反馈优化之后，任务执行成功率至少提升10%；支持80%以上典型异常场景的传播路径可视化；提供标准化的人机协同接口，支持外部干预策略的动态注入。系统具备跨平台兼容性，可集成于多种多智能体框架，为复杂智能系统提供可解释、可干预、可恢复的运行保障能力。

4、产出及交付物：

- 多智能体运行支撑平台原型系统，包含状态追踪、异常分析、处置执行与人机接口模块
- 系统接口文档与部署手册
- 所有交付物均具备可扩展性与工程化落地能力，为后续规模化应用奠定基础
- 发表CCFA类期刊（会议）论文1篇
- 申请发明专利1~2篇

专项二：具身智能

1	前沿探索类	机器人大脑技术栈研究：世界模型技术发展现状与趋势
2	单点突破类	机器人多目标强化学习算法研究
3	单点突破类	基于多模态融合的复杂高动态环境下足式机器人自主导航技术研究
4	单点突破类	面向垂直场景的具身导航方法研究
5	单点突破类	面向巡逻场景的图像及视频异常理解的VLM研究
6	单点突破类	基于深度相机的家庭复杂环境下的机器人SLAM技术

课题类型：前沿探索类

1、研究背景：

- 行业需求驱动

开放环境的“常识缺失”问题：现有机器人系统在非结构化环境中表现出显著的认知局限，例如无法理解“把易腐食品放进冰箱”这类包含时间敏感性和物理常识的指令。

工业场景的预测性维护需求：现代智能制造要求机器人具备前瞻性故障预判能力，如提前识别机械臂关节磨损导致的轨迹偏差。

- 技术瓶颈分析

当前世界模型技术面临几大核心瓶颈：

感知维度缺失：当前的大多数世界模型仅仅引入了视觉，缺乏触觉、听觉、味觉等信息，使智能体无法建立真正的世界概念。

认知断层：现有系统难以建立环境动态变化的mental representation（心智表征），导致对物理交互过程的预测准确率不足60%。

仿真与现实鸿沟：虚拟训练获得的世界模型在真实场景中的物理一致性平均仅有45-68%，特别是在接触力、摩擦系数等物理参数上差异显著。

实时性矛盾：长周期预测(如10分钟级)的计算量级与机器人控制所需的毫秒级响应形成尖锐矛盾。

2、研究内容：

- 动态环境如何构建具身智能模型
- 视觉-语言-动作多模态模型如何对齐，如何提高不同传感器数据（如触觉、听觉传感器）的时空匹配
- 具身智能模型（如世界模型）如何在端侧部署
- 世界模型未来发展趋势预测

3、研究目标：

- 对当前世界模型和VLA技术进行深入调研，为企业决策提供参考。

4、产出及交付物：

- 一篇详细的调研报告。

课题类型：单点突破类

1、研究背景：

近年来，随着强化学习的广泛应用，足式机器人的行走能力取得了显著突破，已能够有效应对各类结构化与非结构化的崎岖路面以及复杂的外部扰动。然而，在工业巡检等实际应用场景中，用户通常期望机器人在确保机身不摔倒的基础上，优先实现对速度和轨迹的高精度跟踪，并在此基础上进一步优化机体的姿态控制，保持平稳运行。然而，当前主流的强化学习方法多采用单一模型架构，在训练过程中多个奖励目标同时作用，容易产生冲突，难以实现对不同目标的灵活切换与协调控制，无法确保多个任务的优先级顺序，限制了机器人在多样化场景下的适应能力。针对传统控制模型的多目标优化策略已有较多研究，例如矢量量化、帕累托优化等，然而，针对强化学习算法的多目标算法还鲜有研究。因此，研究面向多目标的强化学习控制算法，将有助于提升足式机器人在巡检、救援等复杂任务中的自主性与实用性，进一步拓展其应用范围。

2、研究内容：

- 多目标强化学习，研究如何设计和实现能够同时优化多个相互冲突目标的强化学习算法；
- 基于自主决策的层级RL算法，通过融合多传感器信息（视觉、雷达、IMU等），设计位于上层的任务决策器，动态调度最优的底层运动模型；
- 基于目标优先级和帕累托约束的动态奖励函数设计；
- 真机部署验证与仿真-现实迁移技术；

3、研究目标：

- 至少四种地形（平地，草地，山地，楼梯）的足式机器人自适应协同控制
- 平地移动速度跟随误差<5%，转向速度跟随误差<5%
- 平地机体姿态控制误差<1.5°

4、产出及交付物：

- 一套能够在真机验证的完整运动控制系统。
- 机器人顶级会议或SCI期刊论文1篇+1个高质量专利。



课题类型：单点突破类

1、研究背景：

近年来，足式机器人凭借其出色的地形适应性及其仿生外观，在工厂，变电站，车间进行巡检，搬运物料，门店导购等场景中得到逐步应用。同时多模态融合技术结合了视觉的丰富语义信息与激光雷达的高精度几何测量，在自动驾驶领域发挥重要作用。然而，现有的多模态融合方法在足式机器人导航中的实际应用仍存在较大挑战：（1）计算资源开销大，在端侧部署难以满足机器人导航的实时性要求；（2）对动态环境变化响应不足，缺乏高效的时序建模和对动态目标的移动趋势判断，特别是人流密集的动态环境中自主导航；（3）感知与运动规划耦合不足，限制了足式机器人在智能巡逻、门店服务、应急救援等真实场景中的应用落地。因此，有必要构建一种适用于足式机器人的多模态融合感知方法，支持足式机器人在复杂动态环境下的自主决策规划能力。

2、研究内容：

多模态传感器信息对齐与融合方法；面向动态环境的多运动目标跟踪及移动意图预测；结合多模态感知结果实现自主决策与规划；全局/局部路径协同优化策略；针对端侧计算资源受限平台的模型轻量化部署。

3、研究目标：

依托联想Daystar Bot MX(四轮足)和LX(人形)等机器人平台，构建一套基于多模态融合复杂高动态环境下的足式机器人自主导航系统，具体目标包括：

- 实现基于视觉、激光雷达、超声波雷达等传感器数据融合的统一空间环境建模，提高机器人对环境几何与语义的理解能力；
- 支持多目标检测与动态障碍物跟踪及运动预测，提升机器人在复杂场景下的环境适应性；
- 构建与多模态感知深度耦合的自主避障策略与运动规划；
- 在计算资源有限的机器人端侧平台上实现轻量化快速部署，保证系统稳定性与鲁棒性；
- 在智能巡逻、门店服务、应急救援等应用场景中开展实际部署与技术验证。

4、产出及交付物：一套基于多模态融合的复杂高动态环境下（如园区或门店人流密集场景）足式机器人自主导航原型系统；基于联想自研机器人平台进行真机验证；在机器人典型应用场景下的实验验证报告与技术测试数据；SCI论文1篇，专利1~2篇。

课题类型：单点突破类

1、研究背景：

现在落地场景的机器人，基本使用基于SLAM和传统全局规划及局部控制的导航方法，这会受限于定位建图精度和特定场景控制器调参难度。另一方面，我们看到自动驾驶场景的基于BEV或OCC的感知、预测、规划方案已经落地于各家L2+级别的辅助驾驶，具有更好世界通识的VLM/VLA等方案也在一些厂商初步使用。机器人由于其使用范围不局限于结构化场景，同时又具备较强的可通过性，研究机器人在非结构化场景下的智能导航有很强的落地价值。

现有的具身智能导航算法和模型在通用场景下性能距离实际应用落地还有较大距离，但在某些特定垂直场景下（如园区、仓储、家庭）通过结合场景数据和先验来实现具身智能导航具有落地的可能。

2、研究内容：

- 研究点导航/视觉导航/视觉语言导航的不同实现路径，包括强化学习导航、VLA、Diffusion Policy等；
- 在单个垂直场景下，研究基于预训练导航模型的真实数据微调、强化学习微调，或Real2Sim2Real实现路径；
- 研究在一类垂直场景下可泛化的具身导航。

3、研究目标：

- 实现一套在单个场景上点导航成功率达到99%的方法，并在联想机器人平台可验证；
- 实现在一类垂直场景（例如工厂园区、家庭）可泛化的具身导航方法，点目标导航成功率达到95%；
- 在InternScenes数据集上点导航成功率超越SOTA模型。

4、产出及交付物：

- 算法源代码；
- 研究报告；
- 顶刊/顶会论文一篇

课题类型：单点突破类

1、研究背景：

随着人工智能与机器人技术的飞速发展，自主巡逻机器人在安防、工业巡检、园区管理等领域的应用日益广泛。当前，巡逻机器人的核心能力瓶颈已从基础的导航与避障，转向对复杂动态环境的深度认知与理解。传统的基于监督学习的视频异常检测方法，严重依赖大规模、精细标注的异常事件数据集。然而，在真实世界中，异常事件具有类型无限（开放集）、发生频率低、定义模糊等特点，使得提前收集并标注所有可能的异常数据成为一项几乎不可能完成的任务。这些传统方法通常只能识别预先定义的少数几类异常，对于从未见过的、需要常识才能判断的异常情况（例如，“会议室的门在非工作时间被人打开”或“本应在墙上的灭火器被放在了地上”）束手无策，缺乏泛化能力和高级推理能力。

近年来，视觉语言大模型（VLM/MLLM）展现了强大的零样本（Zero-Shot）识别和常识推理能力。它们通过在海量图文数据上的预训练，内化了丰富的世界知识和物理常识，能够像人类一样对视觉场景进行描述、分析和推理。这为解决机器人巡逻领域的开放集异常检测问题提供了一个全新的、从“数据驱动”转向“知识驱动”的革命性范式。

2、研究内容：

在巡逻场景中的非固定点位条件下，研究如何基于VLM实现仅通过设定不同规则提示词或者少量示例完成异常判断，其主体对象主要围绕人员及设备，包括但不限于：人员类：（1）未穿戴/佩戴检测：未戴安全帽、未穿反光衣、未戴口罩、未系安全带等。（2）违规行为检测：在禁烟区吸烟、打电话、玩手机、明火作业等。（3）危险动作检测：攀爬、翻越围栏、高空抛物、打架斗殴、快跑/追逐。

设备类：（1）指示灯状态异常：识别设备的运行、停止、报警等指示灯的颜色和闪烁状态。（2）外观缺陷检测：表面裂纹、锈蚀、漏油、漏水、螺丝松动、线路破损等。（3）环境异常检测：设备周边出现烟雾、火焰、大量粉尘、异物（如工具遗留）。

3、研究目标：

针对机器人巡逻场景，构建一个基于VLM的图像及视频异常理解系统，具体目标包括：

- 设计一套基于VLM的零样本和少样本异常检测方法；
- 实现一个机器人巡逻异常理解原型系统。

4、产出及交付物：

- 算法源代码。
- 研究报告。
- 顶刊/顶会论文一篇。

课题类型：单点突破类

1、研究背景：

家庭消费场景下的小型轮式机器人是目前机器人的热门发展方向，机器人（高度约0.5米，底盘直径约0.3m）需要在家庭环境下实现地图构建，自定位，导航，避障绕行等功能。对于可移动机器人来说，SLAM技术是其核心交互功能，目前基于激光雷达的SLAM方案在业界较为成熟，然而激光SLAM由于要保证大尺度的前向FOV，需要在结构上做较大改动，影响ID设计。而基于深度相机（RGBD）的视觉SLAM方案则可以利用主动深度传感器得到3D深度信息来完成SLAM算法，避免额外的结构影响。在家庭环境下，存在落地窗等透明玻璃，高反射或黑色墙面等场景，由于光的透射和反射作用，深度相机得到的深度信息不准，或无有效深度信息，会直接影响SLAM的算法精度，从而无法闭环。另外，当家庭环境发生较大幅度的变化时（沙发，桌子，柜子等），视觉图像的特征值会无法匹配，从而无法定位。

2、研究内容：

- 从深度相机前端设计上研究更准确、噪声更低的深度深度获取方式，或从后端算法上改善，在有噪声的深度数据前提条件下，提升SLAM算法完成度和准确性。
- 在家庭场景下环境大幅变化后在算法层面上提升重定位的成功率与准确度，并进行地图的局部更新。
- SLAM核心传感器为RGBD深度相机，平台基于RK3588。

3、研究目标：

- 在上述大面积落地窗/高反射墙面/黑色墙面场景下，解决V-SLAM表现差的问题。指标：在同时存在宽度 $\geq 5m$ 落地窗、高反射墙面、黑色障碍物的环境下，完成机器人自定位、自动建图等SLAM功能（机器人行进速度 $\geq 0.5m/s$ ），定位精度 $\pm 10cm$ ，任务达成率 $\geq 99\%$ ，因复杂环境造成的额外时间 $< 50\%$ 。
- 解决家庭环境大幅变化后的重定位问题并进行地图更新。指标：超过10个环境物体变化后，仍能进行机器人自定位，并更新地图，任务达成率 $\geq 99\%$ ，重定位时间 $< 30s$ 。

4、产出及交付物：

- 算法交付物，包括：（1）设计文档，接口文档及头文件（2）基于RK3588平台的算法源代码，二进制安装包及编译环境说明（3）测试用例，测试报告。
- 实现上述目标且可用于家庭服务的机器人硬件方案1套。
- 相关专利交底书1-2篇。
- 相关论文1-2篇。

本项目可根据申报者的合作方案与研究基础，对预期产出与交付物进行相应调整。¹⁸



Lenovo 联想

中国人工智能学会-联想 蓝天科研基金

蓝图汇众智 天高任鸟飞