

# 面向居家养老场景的高质量具身智能仿真数据集构建研究

田佳蕊<sup>1,2</sup>, 孙奥兰<sup>1</sup>, 瞿晓阳<sup>1</sup>, 张旭龙<sup>1</sup>, 付佳<sup>3</sup>, 叶志伟<sup>4</sup>, 王健宗<sup>1</sup>

1. 平安科技(深圳)有限公司, 广东 深圳 518046;

2. 香港中文大学(深圳), 广东 深圳 518172;

3. 湖北省气象工程技术中心, 湖北 武汉 430074;

4. 湖北工业大学, 湖北 武汉 430068

## 摘要

人口老龄化持续深化, 居家养老成为我国主流养老模式。具身智能机器人是破解照护资源短缺、应急响应滞后的关键技术载体。具身智能模型训练依赖大规模高保真交互数据, 而真实场景采集存在安全风险高、成本高、样本不足等问题, 通用仿真数据集未适配养老场景特性、人机安全约束与核心服务任务, 难以直接支撑算法研发与落地。为此, 提出面向居家养老场景的高质量具身智能仿真数据集构建方法, 采用 Marble 世界模型与 Isaac Sim 物理双引擎协同, 搭建场景生成、物理建模、控制执行、数据采集 4 层架构, 分别设计四足机器人应急响应、双足机器人日常照护的专属流程, 融入安全物理约束, 生成多模态、标准化、高泛化的专用仿真数据, 为机器人虚实迁移与产业化落地提供支撑。

## 关键词

具身智能; 物理仿真; 世界模型; 空间智能; 物理智能; 机器人; 视觉语言导航; 视觉语言动作

中图分类号: TP242.6

文献标志码: A

doi: 10.11959/j.issn.2096-0271.BDR26165

## *Research on the construction of high-quality embodied intelligence simulation datasets for residential elderly care scenarios*

Tian Jiarui<sup>1,2</sup>, Sun Aolan<sup>1</sup>, Qu Xiaoyang<sup>1</sup>, Zhang Xulong<sup>1</sup>, Fu Jia<sup>3</sup>, Ye Zhiwei<sup>4</sup>, Wang Jianzong<sup>1</sup>

1. Ping An Technology (Shenzhen) Co., Ltd., Shenzhen 518046, China

2. The Chinese University of Hong Kong, Shenzhen, Shenzhen 518172, China

3. Hubei Meteorological Engineering and Technology Center, Wuhan 430074, China

4. Hubei University of Technology, Wuhan 430068, China

## Abstract

With the intensifying trend of population aging, residential elderly care has become the predominant modality in China. Embodied AI robots are increasingly recognized as a pivotal technical carrier to mitigating the scarcity of caregiving resources and the latency in emergency responses. The training of embodied models necessitates large-scale, high-

fidelity interactive data; however, empirical data acquisition is severely constrained by high safety risks, prohibitive costs, and insufficient sample diversity. Furthermore, general-purpose simulation datasets frequently fail to encapsulate the idiosyncratic environmental features, human-robot safety imperatives, and core service tasks essential to elderly care, thereby hindering industrial deployment. To address these gaps, this paper proposes a systematic methodology for constructing embodied intelligence simulation datasets specifically for residential care scenarios. Leveraging a synergistic dual-engine framework that integrates the Marble World Model with NVIDIA Isaac Sim, we establish a four-tier architecture comprising scene generation, physical modeling, control execution, and data acquisition. We further design specialized operational pipelines for quadrupedal robots in emergency response and bipedal robots in daily assistance, incorporating rigorous physical safety constraints. This approach facilitates the generation of multimodal, standardized, and highly generalizable domain-specific simulation data, effectively filling the data void in the field of embodied intelligence for elderly care and providing robust technical support for sim-to-real transfer and large-scale industrialization.

### Key words

embodied AI, physical simulation, world model, spatial intelligence, physical intelligence, robotics, vision-language navigation (VLN), vision-language-action (VLA)

## 0 引言

具身智能是人工智能与机器人技术深度融合的前沿方向，强调智能体依托物理实体在环境中完成感知、推理、决策与动作执行的闭环交互，是居家养老服务机器人实现自主化、智能化落地的核心技术支撑。高质量交互数据集是训练具身智能模型、优化控制策略、保障部署安全的关键基础。真实数据集虽具备原生物理真实性与行为自然度，但受采集成本高、周期长、危险场景难以复现等因素制约，难以规模化获取。低成本、可扩展、高可控的仿真数据构建已成为当前具身智能研究的主流数据解决方案。

第七次全国人口普查结果显示，我国60岁以上人口达26 400万人，占总人口的18.7%，较第六次人口普查提升5.44个百分点。我国人口老龄化程度持续加深，并进入深度老龄化阶段。在此背景下，居家养老凭借契合本土养老习惯、适配老年生活场景的优势，成为我国主流养老模式，

但照护人力短缺、应急响应滞后等问题日益突出。养老场景中，机器人与老年人、家居环境的交互数据采集面临安全风险高、伦理约束严格、场景个性化强、样本多样性不足等固有障碍，采集的数据无法为跌倒检测响应、物品递送、辅助起身等核心任务提供稳定数据支撑。

为破解真实数据采集受限的难题，本文聚焦居家养老垂直应用场景，在系统梳理具身智能仿真数据集发展现状与技术局限的基础上，提出一套贴合养老场景特性、融入安全物理约束、覆盖核心服务任务的仿真数据集构建方法，旨在填补养老领域具身智能仿真数据缺口，为养老服务机器人的算法研发、策略验证与产业化落地提供可靠数据支撑。

## 1 居家养老具身智能数据集构建背景与发展现状

### 1.1 构建背景

本研究的核心目的在于面向居家养老智能化服务的现实刚需，攻克真实养老场

景下具身智能交互数据采集受限与通用仿真数据集场景适配性不足的双重技术难题，构建一套适配居家养老专属特性、覆盖核心服务任务、融入全物理安全约束的具身智能仿真数据集标准化构建方法。

随着老年人口规模不断扩大，传统居家照护模式面临人力供给不足、应急处置不及时等现实困境，具身智能机器人凭借自主感知、动态决策、物理交互的闭环能力，成为破解居家养老服务瓶颈的关键技术手段<sup>[1]</sup>。

养老服务场景具有高安全需求、多任务覆盖、高人机交互依赖性的核心属性，而居家养老作为养老服务的核心场景之一，呈现出非结构化、多模态、适老化专属与弱标准化的复杂环境特征。从环境维度看，居家养老环境无统一的空间布局与物品摆放规范，障碍物分布不规则，而且需要配备扶手、助行器、护理床等适老化设施，空间紧凑、光照条件复杂，与标准住宅环境存在明显差异。从交互属性看，机器人需与老年群体完成肢体接触、物品递送、应急救援等高风险交互，需严格遵循力反馈、碰撞阈值等物理安全规则，兼顾老年人的生理与心理特征。从信息属性看，场景融合视觉、深度、力觉、语言指令等多源异构数据，而且老年人行为、环境状态具备强动态性，突发风险场景频发，对数据的泛化性与覆盖度提出极高要求。上述特性决定了居家养老具身智能数

据构建无法沿用通用场景的技术方案，必须依托定制化仿真架构与物理约束规则，构建专属化的数据集体系，为养老具身智能的技术研发与实景落地提供可靠的数据基础。

1.2 发展现状

在具身智能研究范式的演进过程中，仿真数据集经历了从静态场景重建到动态任务生成的系统性转变，其发展路径可以概括为5个阶段（如图1所示），即从基础仿真环境的数据构建逐步过渡到世界模型与生成式数据引擎驱动的按需仿真范式。

如图1所示，在基础仿真环境阶段，研究重点集中于真实世界场景的数字化建模与仿真环境构建。代表性数据集有SUN3D<sup>[2]</sup>、Matterport3D<sup>[3]</sup>以及Gibson<sup>[4]</sup>等。该阶段的核心任务在于通过RGB-D扫描与多视角重建技术获取高保真的室内三维结构，并在此基础上构建可用于导航与感知研究的仿真环境。其中，Matterport3D提供了大规模真实室内空间的精细建模，成为后续多个仿真平台的底层场景来源。

在导航与指令数据集阶段，研究范式由场景建模转向语言引导的行为决策。代表性数据集包括R2R<sup>[5]</sup>、CVDN<sup>[6]</sup>以及REVERIE<sup>[7]</sup>。这一阶段的关键创新在于，

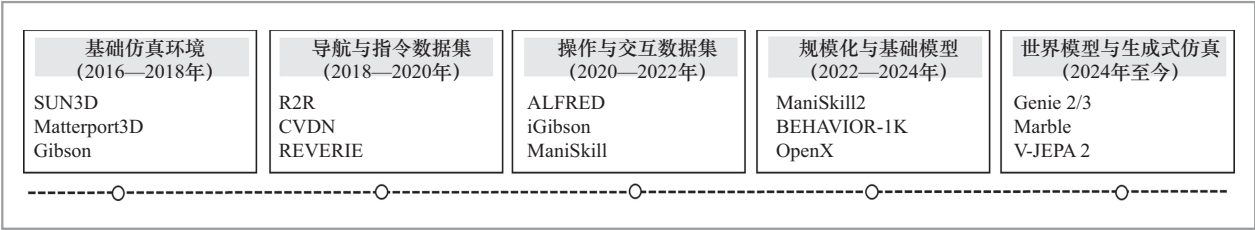


图1 具身智能仿真发展阶段

将自然语言指令与视觉观测及导航轨迹进行强绑定，构建了指令、感知、行动的三元耦合数据结构。

在操作与交互数据集阶段，具身智能研究系统性引入物理交互与长时序任务。代表性数据集包括 ALFRED<sup>[8]</sup>、iGibson<sup>[9]</sup>以及 ManiSkill<sup>[10]</sup>等，该阶段的核心突破在于从导航目标扩展为长时序任务完成。然而，该阶段仍面临显著挑战：数据规模有限、任务泛化能力弱、不同机器人之间缺乏统一表示。这些问题促使研究进入规模化与通用化阶段。

在规模化与基础模型阶段，仿真数据集开始向大规模、多任务、跨机器人本体的方向发展。代表性数据集包括 ManiSkill2<sup>[11]</sup>、BEHAVIOR-1K<sup>[12]</sup>以及 Open X<sup>[13]</sup>。这一阶段的关键特征在于数据规模和任务复杂度的指数级提升，ManiSkill2 扩展了操作任务的多样性与物理精度；BEHAVIOR-1K 将家庭活动扩展至千量级场景与复杂行为组合。与此同时，仿真数据与真实数据的融合成为重要研究方向。然而，该阶段的核心瓶颈逐渐显现：数据获取成本高、长尾任务覆盖不足、数据分布静态且难以扩展。这直接推动了下一阶段向数据生成能力的转型。

在世界模型与生成式仿真阶段，近年来涌现出一批代表性工作并推动该范式快速演进。Genie 2/3 作为可交互的基础世界模型，能够依据单张图像或文本提示以自回归方式生成可操作、时空一致的三维交互环境，为智能体提供按需生成、可持续交互的训练场景。Marble 面向持久化三维世界构建，可由参考图像或文本直接生成结构完整、可自由探索并支持用标准格式导出的三维场景，为下游物理仿真提供高质量、可编辑的场景资产。V-JEPA 2 是基于视频的自监督世界模型，通过联合嵌

入预测在隐空间中学习物理动态规律，具备对环境未来状态的预测能力与零样本机器人规划潜力。上述世界模型推动具身仿真数据从采集与重建向按需生成跃迁，但其生成内容在物理一致性、长时序可控性与仿真/真实数据融合可靠性等方面仍面临突出挑战，这也正是本文采用 Marble 世界模型负责场景生成，并引入 NVIDIA Isaac Sim 物理引擎进行物理约束校验的双引擎协同设计动因。

### 1.3 核心技术挑战

通用数据集在数据规模与多样性方面取得了长足进步，但将其直接应用于居家养老场景仍面临明显的领域鸿沟，具体体现在以下方面。

首先，通用数据集与居家养老领域仿真数据集存在任务语义空间的分布偏移，使得居家养老场景特定任务无法被满足。当前主流数据集涵盖的操作任务以桌面级抓取、物体重排等短时标准化动作为主，其任务语义空间与居家养老场景的核心需求存在根本性差异。养老护理中的典型任务往往具有长时序、多步骤、高度情境依赖等特征，而现有数据集仍缺乏对此类任务的示教与标注信息。

其次，通用数据集与居家养老领域的交互对象存在差异，致使其在居家养老领域内的应用存在安全性挑战。通用操作数据集的交互对象以刚性或半刚性物体为主，偶尔涉及简单的柔性物体操作。然而，居家养老场景中具身智能系统的核心交互对象是老年人的身体，既具有个体差异显著，同时也是相对脆弱的交互对象。

最后，现有通用数据集与仿真平台构建的室内场景多以标准化住宅或实验室环

境为蓝本，未充分考虑适老化居住空间的特殊性。居家养老环境通常具有以下特征：空间布局紧凑，配备护理专用设施，物品摆放高度个性化且缺乏规律性，以及常处于低光照条件下<sup>[14]</sup>。这些特征直接影响机器人的导航策略、感知鲁棒性与操作空间规划，而通用场景数据无法有效覆盖。

## 2 相关工作

当前，具身智能的算法迭代与落地部署高度依赖高质量交互数据支撑，本节围绕具身智能仿真数据集的场景分类、技术演进与生成范式展开系统梳理，深入剖析具身智能仿真数据集的发展现状与局限，为本文居家养老专属仿真数据集构建方法提供研究基础与技术参照。

### 2.1 具身智能仿真数据集相关研究

具身智能作为人工智能领域的前沿方向，致力于实现智能体在物理环境中自主感知、交互决策与动作执行。具身智能在居家养老等场景的落地高度依赖大规模、高保真且场景适配的数据集。真实物理场景的数据采集存在成本高昂、危险场景难以复现、样本多样性不足与定制化程度低等固有瓶颈，仿真数据集凭借场景参数可控、交互安全可复现、任务需求可定制的优势，成为支撑具身智能模型训练、策略优化与安全验证的关键基础设施。

#### （1）场景分类及核心应用载体

从应用空间维度划分，具身智能仿真数据集可分为室外与室内两大核心分支，室内空间因与人类日常生活高度耦合，

成为居家养老等民生领域具身智能技术落地的核心研究载体。依托仿真技术构建室内标准化场景库，可实现环境布局、物体属性与交互任务的精细化调控，有效规避实体实验的安全风险与资源损耗，为多形态机器人的环境感知、运动规划与智能决策提供高质量数据支撑。

如图2所示，本文从场景维度对室内具身智能仿真数据集进行系统化分类，整体按照工业空间、商业与服务空间、居住空间、医疗护理空间的逻辑排布，各场景内部以仿真数据集、机器人配置、核心任务为层级展开，清晰呈现不同领域数据集的技术侧重。

其中，居住空间与医疗护理空间是居家养老具身智能技术的核心落地载体，居住空间直接对应老年日常照护的家庭场景需求，医疗护理空间则为健康监测、康复辅助与上门护理等延伸场景提供仿真支撑，二者形成居家养老健康照护的互补闭环，这也是本文的研究核心。

护理区仿真数据集作为居家养老场景的任务延伸，围绕机器人配置形成了两类典型路径。一类面向固定基座单臂机器人的路径，以Isaac for Healthcare<sup>[15]</sup>为代表，聚焦专业护理场景下的精密辅助操作任务，为机器人的给药、健康监测等基础操作技能提供高保真仿真环境。另一类面向移动底盘单臂机器人的路径以Syn Mediverse<sup>[16]</sup>为核心，支撑移动护理机器人在护理区场景中的环境交互与辅助任务，为居家延伸护理场景下的机器人训练提供场景适配的仿真支撑。

居住空间仿真数据集可进一步划分为通用住宅仿真数据集与适老化住宅仿真数据集，二者基于不同的场景目标形成了差异化的机器人配置与平台载体矩阵。其中，通用住宅仿真数据集面向普通家

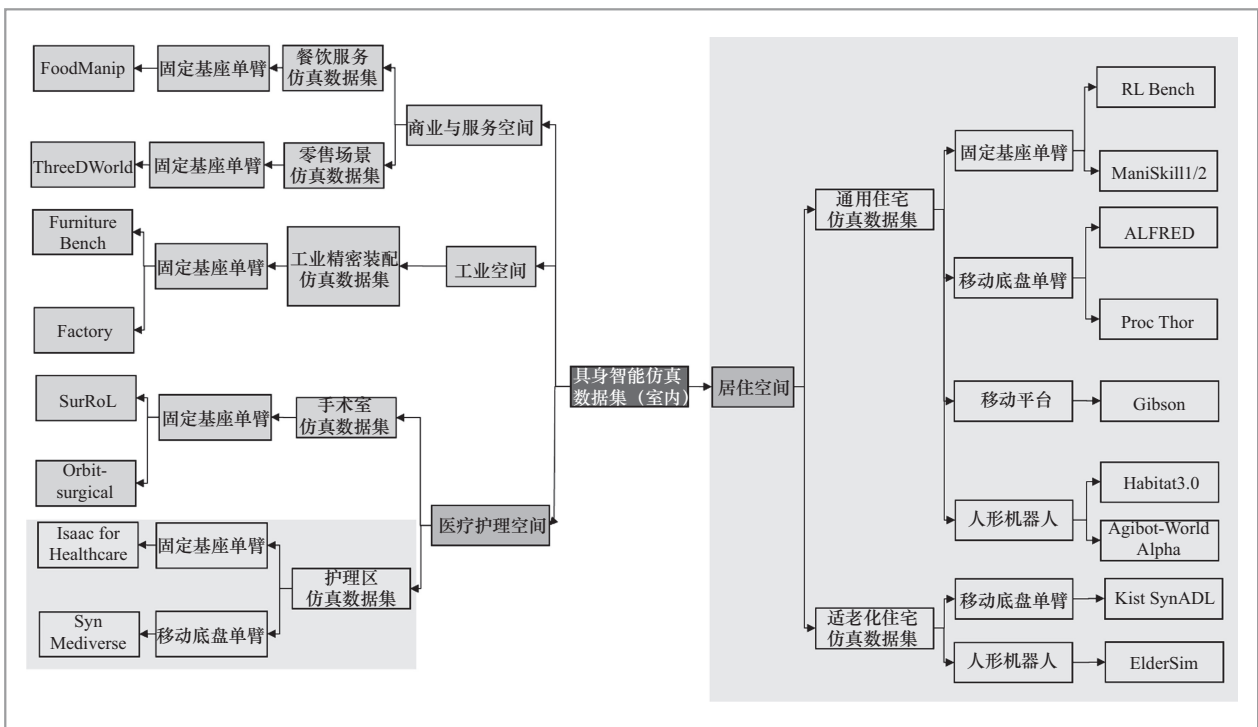


图2 具身智能室内空间仿真数据集分类

庭环境构建，以机器人硬件形态为核心划分维度，形成4类典型技术路径：面向固定基座单臂机器人的路径以RL Bench<sup>[17]</sup>与ManiSkill1/2为核心，聚焦桌面级精密操作任务，为单臂机器人的抓取、装配等基础技能学习提供标准化、可复现的仿真基准；面向移动底盘单臂机器人的路径涵盖ALFRED、Proc Thor<sup>[18]</sup>等，集成家庭场景下的导航、物品交互与语言指令执行任务，支撑移动操作机器人在开放式家庭环境中的泛化能力训练；面向移动平台的路径以Gibson为代表，聚焦家庭场景的视觉导航与环境感知任务，为移动机器人的路径规划与场景理解提供大规模高真实度场景支撑；面向人形机器人的路径则以Habitat3.0与Agibot-World Alpha<sup>[19]</sup>为载体，拓展了家庭场景下全身体态控制、多模态交互与复杂任务协同的训练边界。

在此基础上，适老化住宅仿真数据集专为老年居家场景定制，同样围绕机器人配置形成了两类路径：移动底盘单臂机器人路径以Kist SynADL<sup>[20]</sup>为核心平台，聚焦老年日常活动场景下的辅助照护任务，为适老化场景下的机器人技能训练与策略验证提供定制化支撑；人形机器人路径则以ElderSim<sup>[21]</sup>为核心平台，面向适老化改造家庭环境中的全场景交互任务，为老年照护场景下人形机器人的复杂辅助操作、行为理解与安全交互提供一体化仿真支撑。

从现有体系来看，居住空间虽已形成从通用到适老化、从单臂到人形的多维度覆盖，但适老化住宅分支的平台载体单一，场景多样性与养老专属任务覆盖度仍存在显著缺口，尚未形成针对居家养老场景的、从场景到任务及机器人配置高度耦合的专属仿真数据集体系。

（2）参数对比及特征分析

基于仿真空间类型、细分场景、核心数据集、场景规模、任务复杂度与机器人配置六大维度，进一步对室内具身智能领域主流仿真数据集开展系统性量化对比，以明晰现有数据集的技术边界、硬件适配特性与适老化场景适配短板，为居家养老场景具身智能仿真数据集的构建提供了基准参照。

如表1所示，通用住宅仿真数据集已形成依托机器人硬件配置分层的体系化成熟架构，覆盖固定基座单臂、移动底盘单臂、纯移动平台、双臂/移动双臂4类技术路径，其场景与任务设计高度契合普通家

庭环境特征，但完全未兼顾老年居家环境的适老化专属特性。其中，固定基座单臂机器人路径以RLBench与ManiSkill为典型代表，二者均采用可配置操作场景，聚焦桌面级精密操作任务。移动底盘单臂机器人路径以AI2-THOR、ProcTHOR、Habitat 2.0及ALFRED为核心载体，是家庭具身任务研究的主流支撑平台；纯移动平台路径仅以Gibson为代表，该数据集仅支持纯视觉导航任务、不具备物体操作能力，为移动机器人环境感知研究提供了大规模真实场景支撑。双臂/移动双臂机器人路径由Behavior-1K与Agibot-World Alpha主导，聚焦高复杂度家庭交互任务。

表1 面向居住与医疗护理空间的具身智能仿真数据集参数对照

| 仿真空间   | 类型      | 细分场景 | 核心数据集                                | 场景规模      | 任务复杂度           | 机器人配置  | 核心优势               | 存在局限          |
|--------|---------|------|--------------------------------------|-----------|-----------------|--------|--------------------|---------------|
| 居住空间   | 通用住宅仿真  |      | RL Bench <sup>[17]</sup>             | 可配置操作场景   | 100个分级难度任务      | 固定基座单臂 | 标准基准、任务分级          | 场景单一，无移动      |
|        |         |      | ManiSkill <sup>[10]</sup>            | 可配置操作场景   | 100+个多类型任务      | 固定基座单臂 | 任务多样、场景可配置         | 缺乏长时序任务       |
|        |         |      | Proc Thor <sup>[18]</sup>            | 10 000种场景 | 泛化任务            | 移动底盘单臂 | 程序化生成万级场景、规模大      | 物体交互精度有限      |
|        |         |      | ALFRED <sup>[8]</sup>                | 120+种场景   | 25 000+条指令轨迹    | 移动底盘单臂 | 指令轨迹规模大            | 动作离散化         |
|        |         |      | Gibson <sup>[4]</sup>                | 572个扫描空间  | 泛化任务            | 纯移动平台  | 真实扫描，保真度高          | 仅支持导航，无操作     |
|        |         |      | Behavior-1K <sup>[12]</sup>          | 50个完整住宅   | 1 000+条复杂活动     | 双臂配置   | 涵盖范围广；完整住宅、长时序复杂   | 计算开销大         |
|        | 适老化住宅仿真 |      | Agibot-World Alpha <sup>[19]</sup>   | 100+种场景   | 92 000+条真实机器人轨迹 | 移动双臂   | 真机轨迹规模大            | 采集成本高         |
|        |         |      | ElderSim <sup>[21]</sup>             | 4个背景场景    | 55个任务           | 移动/人形  | 聚焦老年行为与适老环境        | 背景场景稀缺、任务维度单一 |
|        |         |      | KIST SynADL <sup>[20]</sup>          | 5个核心场景    | 55类活动           | 移动底盘单臂 | 可规模化合成             | 缺失物理交互与照护操作   |
|        |         |      |                                      |           |                 |        |                    |               |
| 医疗护理空间 | 护理区仿真   |      | Isaac for Healthcare <sup>[15]</sup> | 可泛化场景     | 泛化任务            | 固定基座单臂 | 高保真护理操作仿真          | 固定基座，无移动      |
|        |         |      | Syn Mediverse <sup>[16]</sup>        | 13种       | 31个语义类别         | 移动底盘单臂 | 多模态语义丰富，支持移动护理场景理解 | 交互任务有限，非居家适老  |

通用住宅数据集虽在场景规模、任务多样性与硬件适配性上已趋于成熟，但其建模对象均为普通青年家庭环境，未覆盖适老化设施、老年行为模式与养老专属服务任务，无法直接支撑居家养老具身智能的研发与落地。

适老化住宅仿真数据集面向居家养老场景专项定制，是本研究的核心对标对象，但其整体仍处于初期探索阶段，存在显著性能短板。ElderSim 构建了 4 类背景场景，覆盖 55 项养老照护任务，支持移动与人形机器人交互，是首个聚焦老年行为与适老化环境的仿真平台；KIST SynADL 基于 5 类核心场景，覆盖 55 类老年日常活动，为移动底盘单臂机器人的养老辅助任务提供数据支撑。上述两类数据集虽实现了养老场景的初步适配，但仍存在场景覆盖不足、老年行为建模粗糙、照护任务维度单一、多智能体交互缺失等核心问题，难以满足养老机器人应急响应、日常照护等全流程任务的训练需求。

### （3）数据集构建方式及其局限性

进一步分析上述数据集的构建方式可以发现，现有研究在数据生产路径上呈现明显的同质化与场景错配。在通用住宅类数据集中，RLBench、ManiSkill 主要依托高保真物理仿真引擎，对桌面场景的物体位姿与物理属性进行程序化随机化，从而批量生成结构化操作数据。ProcTHOR 借助程序化生成技术批量合成万级房屋布局，Habitat 2.0、ALFRED 则在扫描或合成场景之上，通过脚本化策略与自然语言指令标注采集导航与交互轨迹。Gibson 通过真实空间的 RGB-D 扫描重建获取场景，BEHAVIOR-1K 依靠人工定义的家庭活动驱动仿真，Agibot-World Alpha 以人工遥操作真实机器人的方式大规模采集轨迹。面向养老的适老化数据集中，

ElderSim 主要基于预置背景场景与合成的老年人体动作渲染生成，KIST SynADL 通过三维角色动画批量合成老年日常活动数据，二者本质上服务于动作识别等视觉任务。

由此可见，现有构建方式存在三方面共性问题：其一，场景资产多依赖人工预置或真实扫描，生成过程难以低成本、可扩展地覆盖适老化住宅的多样户型与个性化布局；其二，交互数据或来自脚本动作与动作捕捉，或来自纯生成式视频，普遍缺乏高保真物理引擎提供的接触力学真值与人机安全约束，难以支撑老年脆弱交互对象下的安全照护；其三，适老化数据集多聚焦单一动作识别，机器人本体单一，既缺乏应急响应等安全关键任务的闭环采集，也缺乏对照护操作等长时序任务的完整覆盖。相较之下，本文提出的构建方式以 Marble 世界模型驱动场景的自动化、参数化与可扩展生成，突破了人工预置与扫描重建的规模瓶颈；以 NVIDIA Isaac Sim 物理引擎完成质量、摩擦、碰撞等物理属性的精确建模并提供完备真值标注，弥补纯渲染与合成数据在物理属性上的缺失；进一步将力反馈、碰撞阈值等安全物理约束融入动作生成，并面向四足机器人应急响应与双足机器人日常照护分别构建自动化闭环采集流程，实现跨机器人本体、覆盖感知-决策-控制全环节的养老专属数据生产，从而在场景适配性、物理保真度、安全约束与任务完备性上较既有数据集更具优势。

## 2.2 具身智能数据集生成方式

具身智能数据集的生成范式是决定数据物理保真度、可扩展性、标注完备性与跨场景泛化能力的核心前置要素，其底层

技术路径直接决定了具身智能模型的训练收敛效率、交互策略鲁棒性及真实场景部署的落地性能。当前，面向室内具身交互场景的大规模数据集构建技术，已形成表2所示的4类主流生成范式，各类范式在技术实现逻辑、数据生产效率与性能瓶颈上呈现出显著的差异化权衡特征，为不同研究目标下的数据集构建提供了多元路径支撑，各类范式侧重的技术方向存在明确差异。

（1）基于物理仿真的程序化生成

基于物理仿真的程序化生成范式依托高保真物理仿真引擎为载体，核心依托高性能计算能力实现大规模并行处理，通过对场景布局、物体姿态、物理属性等环境变量进行程序化随机化配置，并驱动智能

体自动执行交互任务，实现结构化交互数据的批量化、并行化生产<sup>[30]</sup>。其代表性工作包括面向桌面操作的RLBench与ManiSkill系列仿真平台，该范式凭借GPU集群的并行化仿真能力，具备极高的数据生产吞吐量与场景多样性扩展潜力。同时依托仿真引擎对刚体动力学、碰撞接触模型的精确建模，本范式可提供物体6D位姿、关节力矩、接触力等完备真值标注，为具身智能模型的可复现性训练与大规模并行数据采集提供关键支撑。然而，仿真环境与真实世界间的现实鸿沟始终是该范式的核心局限，接触建模精度不足、高频物理效应模拟偏差等问题，易导致模型在真实机器人部署场景下的泛化性能衰减，难以直接支撑对接触交互精度要求极高的

表2 具身智能数据集主流生成范式多维度特征对比表

| 生成方法         | 核心机制             | 代表性工作                                                | 可扩展性 | 物理保真度 | 主要优势    | 关键局限       |
|--------------|------------------|------------------------------------------------------|------|-------|---------|------------|
| 基于物理仿真的程序化生成 | 仿真随机化执行          | RLBench <sup>[17]</sup><br>ManiSkill <sup>[10]</sup> | 高    | 高     | 完美真值标注  | Sim2Real差距 |
|              |                  |                                                      |      |       | 实验可复现   | 精度有限       |
|              |                  |                                                      |      |       | 大规模并行化  | 泛化能力有限     |
| 基于遥操作的数据采集   | 人类遥操机器人          | RoboCasa <sup>[22]</sup><br>MimicGen <sup>[23]</sup> | 低    | 高     | 捕获人类策略  | 劳动密集       |
|              |                  |                                                      |      |       | 语义质量高   | 吞吐量低       |
|              |                  |                                                      |      |       | 真实动作示范  | 操作差异大      |
| 3D重建与数字孪生    | NeRF、3DGS等技术     | RoboTwin <sup>[24]</sup><br>D3Fields <sup>[25]</sup> | 中    | 中     | 照片级渲染   | 重建耗时       |
|              |                  |                                                      |      |       | 真实世界对应  | 物理属性缺失     |
|              |                  |                                                      |      |       | 安全策略测试  | 交互限制       |
| 生成式模型合成      | 大语言/视觉模型<br>驱动生成 | GenSim <sup>[26]</sup><br>RoboGen <sup>[27]</sup>    | 高    | 中     | 任务多样性   | 需验证管线      |
|              |                  |                                                      |      |       | 最小化人工设计 | 依赖LLM能力    |
|              |                  |                                                      |      |       | 利用世界知识  | 易产生幻觉任务    |
|              | 扩散模型或视频模型        | UniSim <sup>[28]</sup><br>Genie <sup>[29]</sup>      | 高    | 低     | 无需显式仿真器 | 常违反物理定律    |
|              |                  |                                                      |      |       | 规模化预训练  | 无精确状态标注    |
|              |                  |                                                      |      |       | 视觉多样性强  | 可控性有限      |

注：“可扩展性”指在单位人力与算力成本下可生成数据的规模与场景多样性的扩展能力（高＝可低成本生成并自由扩展场景；中＝需一定人工设计或较高算力；低＝高度依赖人工采集）；“物理保真度”指生成数据对真实物理规律的还原程度（高＝由物理引擎提供完备物理真值；中＝具备部分物理属性或与真实场景高度对齐；低＝缺乏显式物理约束）。

居家照护、医疗护理等场景。

## （2）基于遥操作的数据采集

基于遥操作的数据采集范式中，人类操作员借助 VR 或触觉交互设备，对实体或虚拟机器人进行直接控制，核心侧重人机交互的自然性与行为示教的真实性，捕获符合人类行为逻辑的自然交互轨迹与任务策略。此类方式生成的数据天然满足真实物理约束，具备极高的物理保真度与行为合理性，能够捕捉高度自然的人类操作逻辑与细粒度交互语义信息<sup>[31]</sup>，为具身智能模型提供贴近真实任务需求的专家演示数据，尤其适用于居家照护、医疗护理等对交互安全性与行为自然度要求极高的场景。但该范式存在显著的性能短板：数据生产过程依赖人工操作，劳动密集度高、吞吐量极低，难以支撑百万级大规模数据集的构建。同时，不同操作员间的操作风格差异、设备传输延迟、交互指令噪声等问题，易引入数据分布偏差，影响模型训练的稳定性与泛化能力。

## （3）3D 重建与数字孪生

3D 重建与数字孪生范式通过 NeRF、3D Gaussian Splatting 等先进 3D 视觉重建技术，对真实世界室内场景进行高精度扫描与建模，核心侧重计算机图形学中的场景复刻与视觉渲染，构建照片级真实感的虚拟数字副本作为数据生成载体。该范式在场景可扩展性与物理保真度上均处于中等水平，能够实现与真实世界高度对齐的视觉渲染效果，为模型提供高真实度的视觉输入，同时支持智能体安全策略的预部署测试与交互行为验证<sup>[32]</sup>。但其局限性同样显著：3D 重建过程耗时耗力，大规模场景的重建成本高。重建后的虚拟场景往往缺失物体质量、摩擦系数、关节阻尼等精确物理属性，而且交互行为的可复现性与多样性受到重建精度的直接限制，难

以支撑复杂具身交互任务的大规模数据生成。

## （4）生成式模型驱动的数据合成

生成式模型驱动的数据合成范式依托大语言模型、多模态大模型或扩散模型等基础人工智能模型的世界知识与语义理解能力，核心侧重 AI 技术的语义建模与多模态生成，构建了两类差异化的生成路径。其一为大模型与仿真器相结合的半监督生成路径，利用大语言模型自动完成任务描述生成、场景布局规划及奖励函数设计，驱动物理仿真器生成结构化交互数据。该路径在可扩展性上表现中等，物理保真度依赖底层仿真引擎实现，其核心优势在于能够大幅拓展任务多样性，降低人工场景与任务设计成本，同时借助模型的先验知识生成更贴合真实居家照护、医疗护理场景需求的任务分布。但生成场景的物理合理性需额外验证与监管，数据质量高度依赖大语言模型的语义理解与常识推理能力，端到端生成管线的稳定性仍存在显著挑战。其二为直接生成路径，利用扩散模型、生成对抗网络或视频生成模型，直接从噪声或文本指令中合成具身交互的视觉帧与动作轨迹数据，无须依赖显式物理仿真引擎。该路径具备极高的可扩展性，可依托互联网规模预训练数据与并行推理能力实现高效数据生产，而且无需显式仿真器即可生成具备照片级真实感的视觉数据，突破了传统仿真场景的多样性限制。但此类生成数据存在违反物理定律的问题，如物体穿透、非惯性运动、接触交互失真等，缺乏精确状态标注与动作时序一致性保障，而且对任务目标的可控性有限，难以直接应用于居家照护、医疗护理等需要严格物理约束的具身控制任务。

整体而言，上述 4 类生成范式在物理保真度、可扩展性与任务可控性上各有侧

重，不存在普适最优路径；面向居家养老等对安全性与物理保真度要求较高的场景，多范式融合，如将3D重建的真实场景副本与物理仿真引擎的动力学约束相结合，或以大模型常识推理优化生成式范式并由仿真引擎校验物理合理性，正成为发展趋势，也为本研究的数据集构建提供了关键技术参考。

3 居家养老场景具身智能仿真数据集构建方法

综合现有研究与应用需求，居家养老具身智能仿真数据集构建面临三大核心技术挑战：第一，居家养老场景的高保真自动化生成难题，需兼顾真实性、合理性与物理交互完整性；第二，应急响应等安全关键任务的闭环控制与数据采集难题，需实现事件触发、机器人控制、数据标注的全流程自动化；第三，养老交互动作的安全化建模难题，需将力反馈、碰撞约束等物理规则融入动作表征与数据生成过程。

本文面向居家养老这一高度细分且具备强现实需求的垂直应用场景，提出针对性的具身智能仿真数据集构建方法。该方法立足居家环境交互复杂、服务任务专属、机器人本体多样的场景特性，确立以分场景定制、双引擎协同、流程自动化为核心的构建思路，围绕老年群体照护需求设计专属化构建逻辑。整体技术框架划分为场景生成层、物理建模层、控制执行层、数据采集层四大核心层级，形成层级清晰、衔接顺畅的结构化构建体系；同时针对居家场景下应急响应、日常照护等高频核心服务任务，分别设计差异化、精细化的专属构建流程，突破单一任务数据集的应用

局限，并实现跨机器人本体的数据集构建模式，大幅提升数据集的适配范围与实用价值。

3.1 总体构建框架

本文搭建居家养老具身智能仿真数据集，设计一套从虚拟环境生成到多模态数据标准化输出的四层一体化架构。该架构以两大核心居家养老服务任务为顶层需求牵引，所有层级的设计与迭代均围绕养老场景的实际交互逻辑展开，确保数据集与真实服务需求高度匹配。架构采用自底向上的逐层递进式构建模式，底层为上层提供基础支撑，上层为下层明确任务导向，各层级之间通过统一的标准化接口实现数据传输、指令下发与状态反馈的无缝衔接，完整打通数据流与控制流的全链路。依托该标准化架构，能够系统完成虚拟居家场景复刻、物理属性精准建模、机器人交互执行控制、多维度交互数据采集标注等全流程工作，最终生成覆盖机器人环境感知、行为决策、运动控制全环节的高质量居家养老专属仿真数据集，为后续具身智能模型的训练优化、性能验证与真实场景部署提供坚实的数据支撑。图3所示为本方法的核心架构。

(1) 场景生成层

场景生成层是整个数据集构建流程的基础，负责提供高保真、多样化的三维居家环境数据。该层以世界模型 Marble 管线为技术引擎，实现从二维参考图像或文本描述到完整三维居家场景的自动化生成。场景生成层的核心输出包括：符合中国居家养老实际空间特征的户型结构，具备物理真实材质与纹理的家具、家电、辅助器具等室内物体资产，以及适配不同任务需求的场景变体。所有生成的三维场景统一

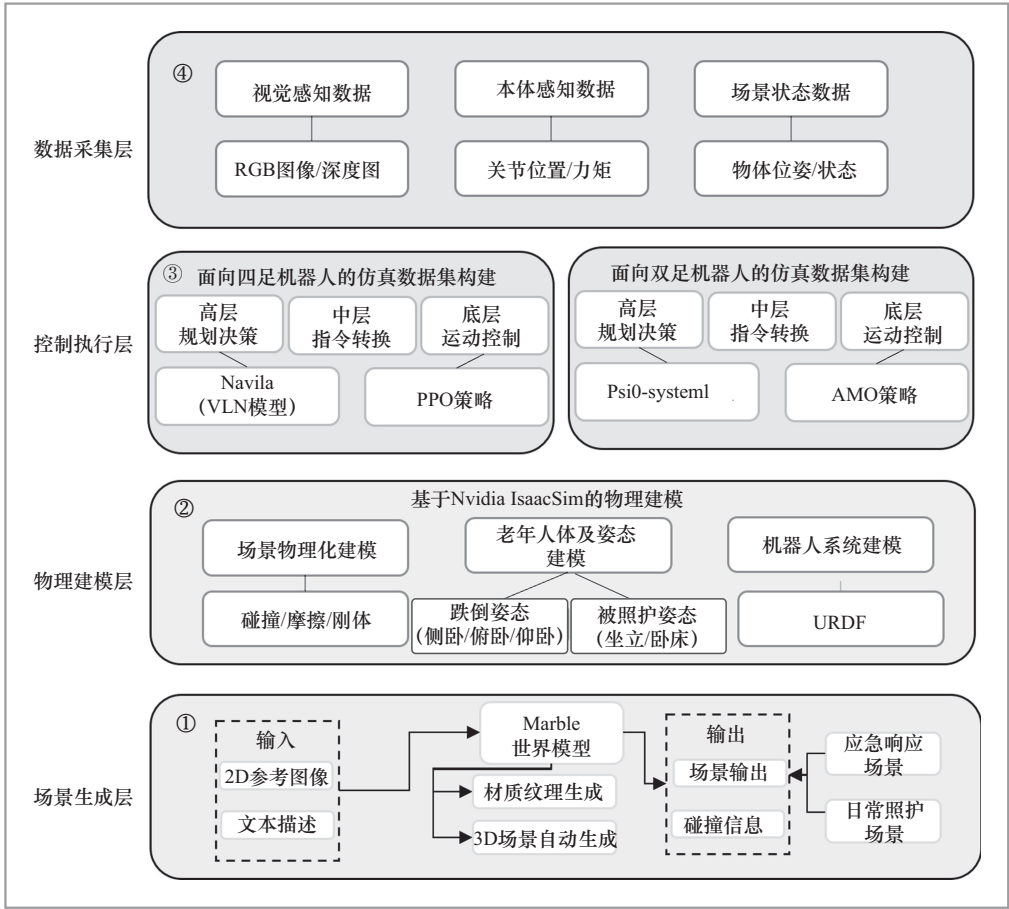


图3 居家养老具身智能仿真数据集构建核心架构

以 OpenUSD 格式输出，确保与下游物理仿真引擎的无缝对接，同时支持场景元素的参数化调控，为后续大规模域随机化数据增广提供基础。

在 Marble 世界模型生成能力的基础上注入适老化空间先验，构建包含扶手、护理床、助行器、防滑垫等适老设施在内的资产库，并将户型紧凑、动线无障碍、低照度等参数纳入生成约束，使生成场景分布贴合中国居家养老真实空间，而非通用青年家庭模板。

### （2）物理建模层

物理建模层承接场景生成层输出的三维场景资产，在 NVIDIA Isaac Sim 物理仿真引擎中完成场景与机器人系统的物理级

精确建模。该层完成场景物理化的核心任务：将 Marble 生成的场景导入 Isaac Sim，为所有物体赋予精确的物理属性，如质量、摩擦系数、弹性系数、碰撞体积等，构建符合真实物理规律的刚体动力学与接触力学仿真环境；特别针对居家养老场景中的高频交互对象进行精细化碰撞网格建模与物理参数标定。同时，该层完成机器人系统建模：根据不同任务需求，导入相应的宇树系列机器人 URDF 模型，完成关节自由度配置、驱动器参数设定、传感器挂载等全系统建模。另外，本层需完成人体模型及姿态建模：实现参数化老年人人体模型，配置符合老年群体的身体尺寸、受限关节活动范围与典型跌倒姿态；同时引入

人机安全物理约束，对机器人与人体的接触设定接触力上限、碰撞阈值与柔顺接触模型，以适配老年人这一脆弱交互对象群体。

(3) 控制执行层

控制执行层是连接仿真环境与数据采集的枢纽，将通用运动、操作策略与养老任务深度绑定，应急响应引入跌倒事件触发机制与安全导航约束，日常照护引入面向人体邻近操作的力控与避障约束。本层负责驱动机器人在物理仿真环境中执行面向特定养老任务的行为策略，采用高层决策规划与底层运动控制的分层控制架构，针对不同任务场景配置差异化的算法组合。控制执行层在仿真环境中通过域随机化技术（如随机化物体位置、光照条件、机器人初始状态、地形参数等），增强所生成数据的多样性与泛化性。

(4) 数据采集层

数据采集层负责从仿真环境运行过程中系统化提取、对齐、标注与存储多模态数据，形成数据集。该层同步录制机器人搭载传感器输出的 RGB 图像、深度图等视觉数据及本体感知数据，覆盖机器人第一人称视角与环境全局监控视角。在通用多模态记录之外扩充养老专属标注字段，包括跌倒姿态类别、救援目标点、人机相

对距离与方位、照护任务语义、接触力等。

3.2 面向四足机器人的应急响应场景数据集构建

为支撑机器人在居家养老场景中的应急响应能力，如图 4 所示，本文基于 Marble 模型生成的高保真 3D 居家场景及机器人系统，构建了多模态、标准化、高适配性的应急响应场景专用数据集，实现应急场景数据从仿真生成到实操验证的全链路闭环，为机器人应急感知、决策规划及运动控制算法的训练与验证提供核心数据支撑。

数据集构建以居家养老高频紧急事件为核心导向，覆盖老年人跌倒等核心应急场景，同时绑定机器人协同响应逻辑。四足机器人负责居家全域移动及环境监测，确保数据集覆盖机器人在应急场景下的感知、运动及交互全维度数据需求，实现数据与机器人控制场景的深度适配。

数据采集采用仿真生成融合模式，依托前期搭建的 Marble-USD 标准化场景与 IsaacLab 仿真引擎，实现数据的规模化、高保真生成。基于 Marble 生成的 3D 场景，通过 OpenUSD 数据交换通道输出高分辨率多模态视觉数据，包括 RGB 图像、深度

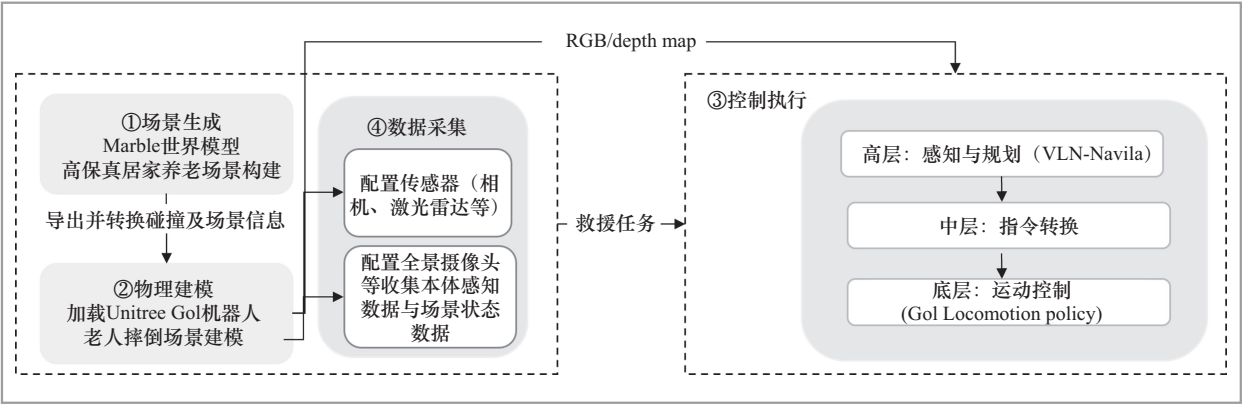


图 4 面向四足机器人的应急响应场景数据集构建流程图(以 UnitreeGo1 为例)

图、语义分割图及 3D 点云数据,覆盖不同光照及场景内条件以及平整、起伏等不同地形的应急场景。

物理建模层完成应急场景内机器人与环境交互系统的构建。在导入的仿真场景中,加载 Unitree Go1 四足机器人本体模型,并完成感知传感器与物理引擎参数的一体化配置。感知层面挂载 RGB 相机与深度相机,输出多模态视觉数据  $D_{\text{Sensor}} = \{I_{\text{RGB}}, I_{\text{depth}}\}$ ,为后续跌倒目标识别与导航提供视觉输入。物理层面标定机器人关节阻尼、刚度与碰撞检测规则,确保仿真环境中机器人的运动学与动力学特性与实物高度一致。构建老年人跌倒应急场景的物理模型,定义跌倒老人的状态与空间特征为  $\mathcal{O}_{\text{elder}} = \{\chi_{\text{elder}}, \sigma_{\text{fall}}, P_{\text{target}}\}$ 。其中,  $\chi_{\text{elder}}$  为老人跌倒姿态下的三维位姿;  $\sigma_{\text{fall}} \in \{0, 1\}$  为跌倒状态标签;  $P_{\text{target}}$  为老人所在目标点,是机器人救援导航的终点。经配置后的物理建模层可接收“前往摔倒老人位置”等救援任务指令,为控制执行层提供可信的机器人与环境交互接口,同时为数据采集层提供多模态感知与状态数据源。

本场景内的控制执行层采用高层感知与规划、中层指令转换、底层运动控制的 3 级架构,实现对老人跌倒场景的精准识别与目标定位导航。高层感知与规划模块基于 Navila 视觉语言导航模型<sup>[33]</sup>,以物理建模层输出的多模态视觉数据与救援任务指令为输入,通过卷积神经网络提取 RGB 与深度图的视觉特征,结合跌倒老人的目标点位与状态,并基于视觉语言对齐机制将自然语言救援指令映射为全局路径点序列  $W$ ,该路径规划过程形式化为最大后验概率估计:

$$W^* = \operatorname{argmax}_W P(W | I_{\text{rgb}}, I_{\text{depth}}, I_{\text{task}}, \tau_{\text{goal}}) \quad (1)$$

中层指令转换模块接收高层输出的路

径点序列与任务约束,将抽象路径规划结果解耦为底层运动控制可执行的标准化指令集:

$$\mathcal{U} = \{v_{\text{ctrl}}(t), \omega_{\text{ctrl}}(t) | t \in [0, T_{\text{nav}}]\} \quad (2)$$

其中,  $v_{\text{ctrl}}(t)$  为  $t$  时刻的期望线速度指令,  $\omega_{\text{ctrl}}(t)$  为期望角速度指令,  $T$  为导航至老人点位的总时长,通过平滑插值与避障优化保证指令的连续性与安全性。底层控制采用基于 Unitree Go1 的 RSL\_RL 框架中的运动策略,通过近端策略优化 (proximal policy optimization, PPO) 算法训练的 Actor-Critic 神经网络策略<sup>[34]</sup>,将中层输出的速度指令转化为机器人 12 个关节的力矩控制信号,实现四足机器人在复杂居家地形中的稳定、敏捷运动。底层运动策略采用独立参数化的 Actor-Critic 双网络架构。策略网络 (即 Actor) 将观测映射为关节动作的概率分布。

本模型通过 3 层全连接神经网络输出动作均值,标准差参数独立可学习。价值网络 (即 Critic) 采用相同网络拓扑但独立参数,输出单一标量表示状态价值估计,用于优势函数计算。训练过程采用同步批量模式,单次迭代可在数千个并行仿真实例中同时执行策略采样,显著提升数据收集效率。

数据采集层在每个仿真时步同步录制异质数据流,并通过仿真引擎提供的全局时钟实现时间对齐。视觉感知数据流以 20 Hz 频率从机器人头部挂载的 RGB 相机与深度相机录制第一视角图像数据,同时配置环境全局俯瞰视角监控相机,以相同频率录制场景鸟瞰画面,用于数据标注校验与算法行为可视化分析。本体感知数据流以 50 Hz 控制频率同步记录机器人完整本体状态向量,包括 12 个关节的位置、速度、力矩测量值,以及 IMU 传感器输出的基座三轴角速

度与加速度。场景状态数据流记录仿真环境中所有动态物体的时变状态，核心包括跌倒老人的三维位姿即位置与朝向、跌倒状态标签（包含侧卧、仰卧或俯卧），以及机器人基座与目标老人之间的欧氏距离与相对方位角。

### 3.3 面向双足机器人的日常照护场景数据集构建

在构建日常照护场景数据集的过程中，以居家养老环境中高频出现的日常生活辅助需求为核心导向，设置面向老年人居家生活的饮食照料、药品管理、物品整理与递送等典型照护任务。数据集构建流程如图5所示。

日常照护场景的三维环境生成层同样基于 Marble 管线自动化生成，重点构建功能性操作空间密集的居家环境，包括厨房区域、卧室与客厅区域等。在场景生成过程中，特别关注可操作物体的精细化建模，每个可抓取物体均配置准确的几何尺寸、质量、质心位置、表面摩擦系数及抓取关键点标注，确保后续机器人操作交互的物理真实性。

生成的照护场景以 OpenUSD 格式导入 Isaac Sim 仿真平台，完成全场景物理属性

配置。针对日常照护场景的特殊需求，物理建模层重点进行可交互物体的刚体动力学建模，包括精确碰撞网格生成、摩擦锥参数标定、物体间接触刚度设定等，确保抓取、放置、推拉等操作交互的物理仿真真实性。在物理建模层需完成 Unitree G1 人形机器人系统建模。导入 G1 机器人 URDF 模型，配置全身 23 个自由度的关节驱动参数，挂载头部 RGB-D 相机，用于场景感知与物体检测。同时挂载腕部 RGB 相机，用于近距离操作引导及抓取力感知的力矩传感器、全身关节编码器等传感器系统。

本场景的控制执行层采用  $\Psi_0$  (Psi-Zero) 端到端学习框架与大语言模型驱动的任务规划相结合的架构<sup>[35]</sup>，实现从自然语言照护指令到机器人全身协调操作的端到端执行。 $\Psi_0$  作为专门面向人形机器人的通用控制策略，能够统一处理移动与操作的耦合控制，在日常照护场景中完成抓取、搬运、放置等精细化灵巧操作。 $\Psi_0$  架构遵循 VLM 骨干网、动作专家、下肢控制器的分层设计理念，3 个系统各司其职且协同联动：System-2 (VLM 骨干网) 负责高层感知与语义理解，System-1 (动作专家) 负责全身动作生成，System-0 (下肢控制器) 负责底层移动稳定性保障。这

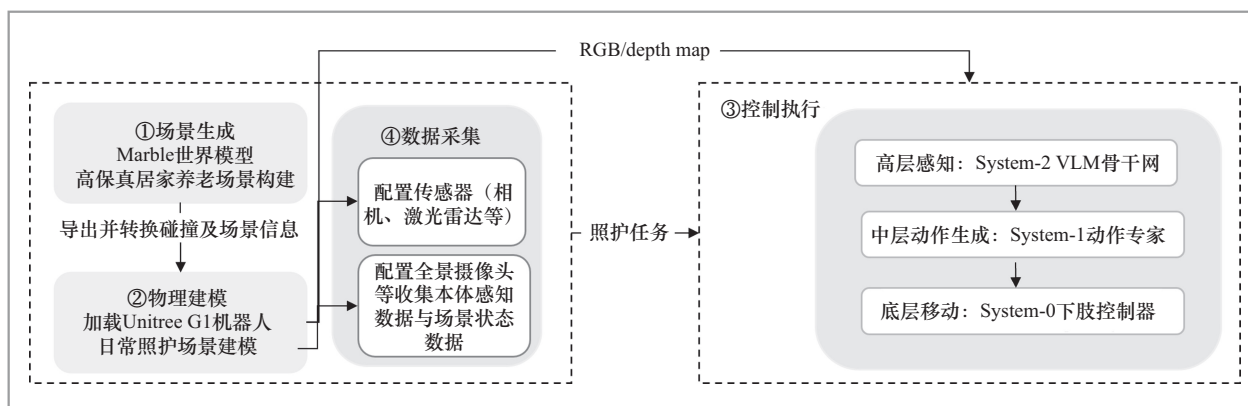


图5 面向双足机器人的日常照护场景数据集构建流程图(以 Unitree G1 为例)

一架构适用于居家照护场景中涉及的复杂移动操作任务，能够在保证上肢精细操作的同时维持全身平衡与稳定移动。System-2是 $\Psi_0$ 的最高层决策模块，基于Qwen3-VL-2B-Instruct预训练视觉语言模型构建，承担场景理解、物体识别与任务语义解析的核心功能。该模块接收来自物理建模层的多视角视觉观测 $J_c = \{J_c^{\text{head}}, J_c^{\text{wrist}_L}, J_c^{\text{wrist}_R}\}$ 与照护任务自然语言指令 $T_{\text{care}}$ ，输出统一的多模态特征表示 $Z_c^{\text{vlm}}$ ，为下游动作生成提供语义丰富的输入。头部RGB-D相机输出的RGB图像与双腕部RGB相机输出的图像被划分为固定大小的图像块，经线性嵌入与位置编码后送入多层Transformer编码器。视觉编码器继承Qwen3-VL在大规模图像、文本对上的预训练权重，具备开放世界物体识别能力，能够准确识别居家环境中的各类日常物品（药瓶、水杯、餐具等）。语言编码过程将照护指令进行词元化并通过Qwen3-VL的语言编码器生成文本特征：

$$Z_c^{\text{lang}} = \text{TextEncoder}_{\text{Qwen}}(T_{\text{care}}) \quad (3)$$

System-1是 $\Psi_0$ 的中层执行模块，采用多模态扩散Transformer（MM-DiT）架构，负责将System-2输出的多模态特征映射为机器人全身动作序列。System-0是 $\Psi_0$ 的底层执行模块，基于AMO（adaptive model-based optimization）强化学习跟踪策略构建<sup>[36]</sup>，负责将System-1输出的8自由度移动指令映射为15自由度的下肢关节角度，保障人形机器人在执行上肢操作任务时的全身平衡与移动稳定性。

数据采集层在仿真运行过程中同步录制异质模态数据流，并通过Isaac Sim全局仿真时钟实现严格的时间戳对齐，形成标准化的照护任务数据集。针对 $\Psi_0$ 三级架构的特点，数据采集层需同步记录各级系

统的输入输出状态，以支撑模型训练、推理分析及系统级调试。视觉感知数据流以10 Hz频率从G1机器人挂载的多视角相机系统录制图像数据。本体感知数据流以100 Hz控制频率录制G1机器人完整本体状态，包含关节状态、IMU数据与末端执行器状态及足端状态。

经过上述流程构建日常照护场景后，双足机器人可在居家场景中执行日常照护任务。数据集涵盖多样化的照护任务指令、高保真的多视角视觉观测、精确的机器人全身状态与动作序列、完整的 $\Psi_0$ 三级系统内部状态记录，以及详尽的物体交互状态。

### 3.4 数据集评价

本文重点聚焦居家养老仿真数据集整套方法论的设计与架构搭建，现阶段完成整体技术方案、4层分层框架、双引擎协同逻辑、两类机器人任务流水线的体系化设计，由于现实中存在诸多安全性风险挑战，尚未完成大批量数据集量产。因此本节从理论定性对比与架构设计合理性两方面开展方法有效性论证。

进一步对标现有适老化数据集的实现思路，本节从设计逻辑层面剖析本文方案优势。首先，从场景生成层面而言，现有适老化数据集场景依赖人工建模，场景拓展成本高。本文依托Marble世界模型实现文本与图像驱动的自动化参数化场景生成，从源头解决养老场景样本单一的痛点。其次，在物理约束层面，现有方案大多仅进行人体动作动画渲染，缺少人机接触力学、老人防护阈值等安全物理规则，本研究使用Isaac Sim完成全物体物理参数标定，将人机安全限值嵌入建模环节，契合养老近距离交互的安全刚需。第三，就任务设计层面而言，当前适老化数据集

缺少机器人闭环交互流程。本文区分四足应急救援、双足日常照护两条独立流水线，面向养老真实业务任务设计控制链路，实现从场景、建模、控制到数据采集的全闭环。

综上，从顶层设计来看，本文构建思路针对性地弥补了现有养老数据集三大短板，在场景拓展性、物理真实性、任务实用性上具备优势。同时，本文整套生成链路基于 OpenUSD 通用工业格式完成数据互通，生成的数据格式兼容主流人形、四足机器人仿真开发框架与实体机器人驱动接口，后续接入实体样机可开展虚实迁移测试。

### 3.5 研究展望

由于现实中存在的安全性风险挑战，本研究及当前居家养老具身智能仿真数据集相关研究仍存在一定局限性，无法将仿真训练后的机器人直接应用于真实养老场景进行大规模验证，这导致虚实迁移的关键指标无法得到真实场景的充分验证。

未来，笔者将从场景任务泛化、虚实迁移校准、多机协同数据、人机安全建模、多模态标准化、产业生态落地等多维度开展系统性深耕，全面提升数据集的完备性、实用性与产业化适配能力。

第一，深化仿真到真实世界的迁移校准研究，基于真实四足机器人与双足机器人在实景居家环境中采集的视觉、力觉、运动轨迹与交互数据，构建自动化虚实参数校准工作流，结合领域自适应算法缩小仿真与现实的分布差异<sup>[37]</sup>，量化验证应急响应时延、操作成功率、导航精度等虚实迁移关键指标。

第二，构建多机器人协同照护专用数据集，设计四足机器人全域巡检、应急定

位与双足机器人精细操作、近身照护的协同任务范式，采集多机感知共享、任务动态分配、交互避障、通信同步等全流程数据<sup>[38]</sup>，制定多智能体协同数据的标注规则与评测基准，支撑多机协作养老服务算法研发。同时，推动数据集开源共享与产业落地，开放核心数据子集与仿真部署工具链，提供基线算法与数据加载接口，联合养老服务机构、机器人企业开展实景验证，形成居家养老具身智能数据集的行业应用规范，加速技术从实验室走向规模化养老服务落地。

第三，持续拓展居家养老场景与任务的覆盖广度，面向半失能、失能、高龄独居等不同健康状况的老年群体，新增居家康复训练、慢病健康监测、认知障碍陪伴、跌倒后二次救援、精准喂药与进食辅助、适老化环境智能调控等细分任务，同时覆盖小户型、适老化改造住宅、社区养老居家等多种户型形态<sup>[39]</sup>，通过参数化域随机化生成海量场景变体，补齐长尾风险场景的数据短板。

## 4 结束语

居家养老场景的非结构化特征、高安全交互需求与专属服务任务，决定了其具身智能数据构建需突破通用数据集的适配性局限，而仿真数据集凭借生成成本低、场景可控性高等优势，成为破解养老领域具身智能数据资源短缺的核心路径。本文针对居家养老具身智能仿真数据集构建的三大核心技术挑战，提出分场景定制、双引擎协同的构建方法，通过场景生成层、物理建模层、控制执行层、数据采集层这4层标准化架构，实现了从高保真三维居家场景生成到多模态交互数据输出的全流

程标准化构建。该方法以 Marble 世界模型与 NVIDIA Isaac Sim 物理仿真引擎为核心技术支撑,完成了贴合居家养老空间特征的场景建模、适配老年人生理特征的人机交互建模,以及针对紧急响应、日常照护等核心任务的差异化控制架构设计,构建的专用数据集实现了跨机器人本体及机器人感知、决策、控制多维度数据的覆盖,填补了居家养老领域具身智能仿真数据的空白,也弥补了通用具身智能数据集在养老场景特征、任务设定、人机安全交互约束上的适配性缺陷。

后续可进一步拓展场景与任务覆盖范围,形成全场景、全任务的居家养老具身智能仿真数据体系;深化 Sim2Real 迁移技术研究,结合真实居家场景的实测数据对仿真数据集进行优化与修正,通过虚实融合的方式提升模型从仿真到真实场景的迁移性能;同时开展多机器人协同养老的数据集构建研究,适配多机协作的养老服务需求。

## 参考文献:

- [1] 丁贵广,朱晨,王潇婉,等.具身智能数据采集与处理综述[J].数据采集与处理,2026,41(2):332-346.  
Ding G G, Zhu C, Wang X W, et al. A survey of datasets collection and processing for embodied intelligence[J]. Journal of Data Acquisition and Processing, 2026, 41(2): 332-346.
- [2] Xiao J X, Owens A, Torralba A. SUN3D: a database of big spaces reconstructed using SfM and object labels[C]//Proceedings of the 2013 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2013: 1625-1632.
- [3] Chang A, Dai A, Funkhouser T, et al. Matterport3D: learning from RGB-D data in indoor environments[PP]. arXiv (2017-09-18) [2026-07-27]. arXiv: 1709.06158.
- [4] Xia F, Zamir A R, He Z Y, et al. Gibson env: real-world perception for embodied agents[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 9068-9079.
- [5] Savva M, Kadian A, Maksymets O, et al. Habitat: a platform for embodied AI research[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2019: 9338-9346.
- [6] Anderson P, Wu Q, Teney D, et al. Vision-and-language navigation: interpreting visually-grounded navigation instructions in real environments[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 3674-3683.
- [7] Qi Y K, Wu Q, Anderson P, et al. REVERIE: remote embodied visual referring expression in real indoor environments[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 9979-9988.
- [8] Shridhar M, Thomason J, Gordon D, et al. ALFRED: a benchmark for interpreting grounded instructions for everyday tasks[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 10737-10746.
- [9] Xia F, Shen W B, Li C S, et al. Interactive Gibson benchmark: a benchmark

- for interactive navigation in cluttered environments[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 713–720.
- [10] Xiang F, Qin Y, Zhuang H, et al. Mani-Skill: generalizable manipulation skill learning from demonstrations[C]//2021 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2021: 6022–6029.
- [11] Gu J Y, Xiang F B, Li X L, et al. Mani-Skill2: a unified benchmark for generalizable manipulation skills[C]//Proceedings of the International Conference on Learning Representations. Kigali: ICLR, 2023.
- [12] Clegg A, Lopez A, Szot A, et al. BEHAVIOR-1K: A large-scale dataset for embodied AI in household environments[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 21104–21114.
- [13] Collaboration O X, Padalkar A, Pooley A, et al. Open X-embodiment: robotic learning datasets and RT-X models[C]//Proceedings of the Conference on Robot Learning. Atlanta: CoRL, 2023.
- [14] 赵博涛, 亢祖衡, 瞿晓阳, 等. 基于多模态大模型的具身智能体研究进展与展望[J]. 大数据, 2025, 11(3): 108–138.
- Zhao B T, Kang Z H, Qu X Y, et al. Review and emerging trends of embodied agent based on multimodal large language models[J]. Big Data Research, 2025, 11(3): 108–138.
- [15] Toloui M, Diaz-Pinto A, Moghani M, et al. NVIDIA Isaac for healthcare: an AI-powered robotics development platform for healthcare[C]//Proceedings of the 2025 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2025.
- [16] Mohan R, Arce J, Mokhtar S, et al. Syn-mediverse: a multimodal synthetic dataset for intelligent scene understanding of healthcare facilities[J]. IEEE Robotics and Automation Letters, 2024, 9(8): 7094–7101.
- [17] James S, Ma Z C, Arrojo D R, et al. RL Bench: the robot learning benchmark & learning environment[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 3019–3026.
- [18] Deitke M, Vanderbilt E, Herrasti A, et al. ProcTHOR: large-scale embodied AI using procedural generation[C]//Proceedings of the Advances in Neural Information Processing Systems 35. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2022: 5982–5994.
- [19] AgiBot-World-Contributors, Bu Q W, Cai J S, et al. AgiBot world colosseo: a large-scale manipulation platform for scalable and intelligent embodied systems[PP]. V4. arXiv (2025-08-04)[2026-07-27]. arXiv:2503.06669.
- [20] Hwang H, Jang C, Park G, et al. KIST SynADL: a large-scale synthetic dataset of elderly activities of daily living[J]. IEEE Access, 2021, 9: 20336–20349.
- [21] Hwang H, Jang C, Park G, et al. Elder-Sim: a synthetic data generation platform for human action recognition in eldercare applications[J]. IEEE Access, 2023, 11: 9279–9294.
- [22] Nasiriany S, Maddukuri A, Zhang L C, et al. RoboCasa: large-scale simulation of everyday tasks for generalist robots[C]//Proceedings of Robotics: Science and Systems. Delft: Robotics Proceedings, 2024.
- [23] Mandlekar A, Nasiriany S, Wen B W, et al. MimicGen: a data generation system for scalable robot learning using human

- demonstrations[C]//Proceedings of the 7th Conference on Robot Learning. London: PMLR, 2023: 1820–1864.
- [24] Mu Y, Chen T X, Chen Z X, et al. Ro-boTwin: dual-arm robot benchmark with generative digital twins[C]//Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2025: 27649–27660.
- [25] Wang Y X, Li Z R, Zhang M T, et al. D3Fields: dynamic 3D descriptor fields for zero-shot generalizable robotic manipulation[C]//Proceedings of the Conference on Robot Learning (CoRL). London: PMLR, 2024.
- [26] Wang L R, Ling Y Y, Yuan Z C, et al. GenSim: generating robotic simulation tasks *via* large language models[C]//Proceedings of the International Conference on Learning Representations. Vienna: PMLR, 2024.
- [27] Wang Y F, Xian Z, Chen F, et al. Robo-Gen: towards unleashing infinite data for automated robot learning *via* generative simulation[C]//Proceedings of the International Conference on Machine Learning. Vienna: PMLR, 2024.
- [28] Yang Z, Chen Y, Wang J K, et al. Uni-Sim: a neural closed-loop sensor simulator[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 1389–1399.
- [29] Parker-Holder J, Fruchter S. Genie 3: A new frontier for world models[J]. Google DeepMind Blog, 2025, 4.
- [30] Zhang Y H, Huang Z Y, Ye L Q, et al. Leveraging VR robot games to facilitate data collection for embodied intelligence tasks[C]//Proceedings of the 2025 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2026.
- [31] Zhang H, Shen Y Q, Soberanis Mu R D, et al. TwinOR: photorealistic digital twins of dynamic operating rooms for embodied AI research[PP]. V2. arXiv (2026-04-16) [2026-07-27]. arXiv: 2511.07412.
- [32] Bilge A, Akyol E, Kuscü M. Reservoir computing-based detection for molecular communications[PP]. arXiv (2025-11-11)[2026-07-27]. arXiv:2511.08762.
- [33] Cheng A C, Ji Y D, Yang Z J, et al. NaVILA: legged robot vision-language-action model for navigation[C]//Proceedings of the Robotics: Science and Systems XXI. Robotics: Science and Systems Foundation, 2025.
- [34] Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms[PP]. V2. arXiv (2017-08-28) [2026-07-27]. arXiv:707.06347.
- [35] Wei S L, Jing H Y, Li B Q, et al.  $\Psi_0$ : an open foundation model towards universal humanoid loco-manipulation[PP]. arXiv (2026-03-12)[2026-07-27]. arXiv: 603.12263.
- [36] Li J L, Cheng X X, Huang T S, et al. AMO: adaptive motion optimization for hyper-dexterous humanoid whole-body control[PP]. arXiv (2025-05-06) [2026-07-27]. arXiv:2505.03738.
- [37] 杨峻峰, 田国会, 邵旭阳. 基于在线信息体构建与维护的长期自治家庭服务机器人环境理解方法[J]. 机器人, 2025(2): 269–282.
- Yang J F, Tian G H, Shao X Y. Environment understanding method for long-term autonomous home service robot based on online information entity construction and maintenance[J]. Robot, 2025(2): 269–282.
- [38] 刘阳, 柏永杰, 林惊. 面向人机电物高效融合与

协作的具身智能技术体系[J]. 机器人, 2025, 47(4): 559-580.

Liu Y, Bai Y J, Lin L. Technological system for embodied intelligence towards efficient integration and coordination of human, robot, and physical world[J]. Robot, 2025, 47(4): 559-580.

[39] 林谦, 余超, 伍夏威, 等. 面向机器人系统的虚实迁移强化学习综述[J]. 软件学报, 2024, 35(2): 711-738.

Lin Q, Yu C, Wu X W, et al. Survey on sim-to-real transfer reinforcement learning in robot systems[J]. Journal of Software, 2024, 35(2): 711-738.

作者简介



田佳蕊（2002-），女，平安科技（深圳）有限公司算法助理工程师，香港中文大学（深圳）硕士生，主要研究方向为具身智能、机器学习。



孙奥兰（1996-），女，平安科技（深圳）有限公司高级算法工程师，主要研究方向为具身智能、信号处理、语音生成、大模型，发表10篇国内外核心论文，获得59项专利授权。



瞿晓阳（1988-），男，博士，平安科技前沿机器学习算法分组负责人，清华大学深圳国际研究生院校外导师，中国科技大学先进技术研究院校外导师，中佛罗里达大学访问学者，长期从事机器学习、大数据、体系结构方面的研究工作，在语音语义分析、自动化机器学习、零样本和小样本学习、高性能计算与存储等方面经验丰富。近几年，在体系结构方向和人工智能方向等国际顶级会议和顶级期刊（INFOCOM、DAC、IPDPS、NeurIPS、IJCAI、ICASSP、Interspeech、TPDS等）发表过近50篇文章，其中1篇论文荣获会议最佳学生论文奖提名，担任多个国际顶级期刊的评委，已授权的专利70篇，已出版的专著2本。



张旭龙（1988-），男，博士，平安科技（深圳）有限公司高级算法研究员，2023年入选上海市东方英才计划青年项目。兼任清华大学深圳研究院及中国科学技术大学先进技术研究院校外导师、中国自动化学会联邦数据与联邦智能专业委员会委员，IEEE、中国自动化学会、及中国计算机学会会员。主要研究方向为大模型、具身智能、跨模态智能计算以及模型上下文协议等。发表学术论文80余篇，获得国家发明专利授权20余项。



付佳（1981-），男，高级工程师，湖北省气象工程技术中心（华中区域人工影响天气技术中心）人影作业科科长。曾获“省气象科技人才杰出青年奖”，两次获“湖北省气象部门青年新秀”称号。主持厅局级以上科研业务项目3项（含省自然科学基金重点项目1项），参与科研业务项目17项，并自主开发了十余项专业气象服务系统。在国内外重要学术刊物及会议上发表论文17篇，其中第一作者8篇（含核心期刊5篇），参与编写专著2本，获发明专利2项。目前主要从事气象数据处理、飞机作业管理等方向的科研与技术应用工作。



**叶志伟**（1978–），男，博士，教授，湖北工业大学计算机科学与人工智能学院院长/大数据与人工智能产业学院（湖北省现代产业学院）执行院长、博士生导师，制造工业数智化技术与应用湖北省工程研究中心主任，全国高等院校计算机基础教育研究会数据科学专委会副主任委员，数据挖掘与数据仓库国家一流课程负责人。在国内外重要学术刊物和会议上发表科研论文150余篇，其中有100多篇论文被SCI或EI收录，总引用次数2 000多次。主持和参与完成国家自然科学基金项目、湖北省重点研发计划等项目10余项，目前主要从事机器学习、图像处理、智能计算等方向的教学和科研工作。



**王健宗**（1983–），男，博士，正高级工程师，平安科技（深圳）有限公司智能金融前沿技术研究院院长。美国佛罗里达大学人工智能博士后，美国莱斯大学和华中科技大学联合培养博士，中国计算机学会杰出会员，中国计算机学会大数据专家委员会委员，中国自动化学会联邦数据和联邦智能专业委员会副主任。主要研究方向为具身智能、大模型、联邦学习和深度学习等。荣获中国专利优秀奖、广东省科技进步二等奖、吴文俊人工智能科技进步奖一等奖、人民银行金融发展一等奖、深圳高层次专业人才领军人才、福田英才等奖项。

录用日期: 2026-07-27  
通信作者: 王健宗, jzwang@188.com  
基金项目: 深圳市重点产业研发计划(No. ZDCY20250901100302003)  
Foundation Item: Shenzhen Key Industry Research and Development Program under Grant(No. ZDCY20250901100302003)