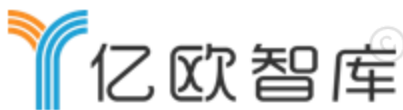


# 2026中国具身智能数据采集与数据产业发展展望

亿欧智库 <https://www.iyiou.com/research>

Copyright reserved to EO Intelligence, June, 2026



用第三方视角和专业服务  
助力产业科技升级和价值创造

- ✓ 研究领域：覆盖人工智能、未来产业、汽车出行、大健康、消费生活、智能制造、电商零售、数字农业、智慧城市、金融科技、物流供应链、企业服务、双碳等多行业领域
- ✓ 服务对象：包含国家部委、地方政府、央国企、互联网科技公司以及外资500强和民营500强
- ✓ 独创模型：亿数合创团队在10余年产业研究和咨询经验的基础上联合科研单位，研发了诊断企业数字化和创新力水平的TOIPO模型。模型从5大维度，30个细分维度对企业的战略、产品、技术、供应链、经营等方面进行全面诊断。

## 亿欧智库历史服务项目

累计发布自研型研究报告600+

定制型研究与白皮书项目300+

战略规划型项目100+



## 亿欧智库服务项目类型



获取更多维度报告数据，请访问亿欧网([www.iyiou.com](http://www.iyiou.com))

## 目录

## CONTENTS

### 01 中国具身智能数采与数据产业发展背景

- 1.1 中国具身智能数据行业发展背景
- 1.2 中国具身智能数采与数据市场规模与驱动因素
- 1.3 中国具身智能数采与数据行业的核心挑战与瓶颈

### 02 具身智能数据采集技术路线对比

- 2.1 遥操作：数据质量最高，成本最大
- 2.2 EGO 第一人称数据：成本较低，规模化潜力最大
- 2.3 便携 UMI：成本与质量的最佳平衡点
- 2.4 仿真数据：产能无限，但 Sim2Real 鸿沟难逾

### 03 具身智能数采与数据行业产业链结构

- 3.1 底层基础设施层
- 3.2 数据采集层
- 3.3 数据加工层
- 3.4 数据应用层
- 3.5 优秀企业案例分析
- 3.6 数据采集案例 TOP20
- 3.7 产业图谱

### 04 具身智能产业商业化突围策略与未来展望

- 4.1 商业模式演进与盈利路径
- 4.2 未来趋势
- 4.3 战略建议

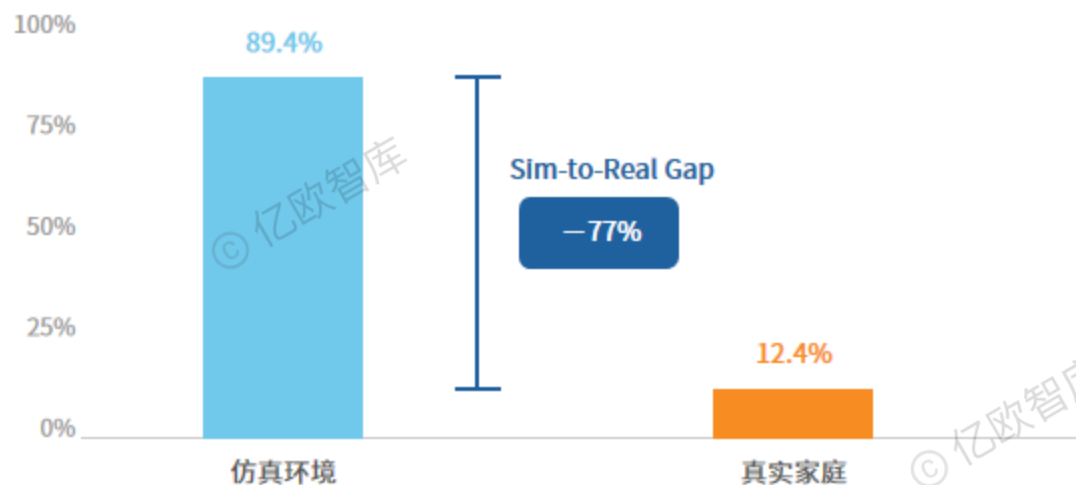
# 1.1 中国具身智能数据行业发展背景

◆ 2026年，具身智能产业正经历一场竞争焦点的结构性转移。过去三年，行业将大量资本和人才投入到硬件本体与模型算法——谁能做出更灵活、更稳定、更具展示效果的机器人演示，谁就能占据资本与舆论的制高点。然而，随着VLA（视觉-语言-动作）模型架构的收敛和硬件供应链的初步成熟，一道无法回避的裂缝正在被撕开，即制约机器人规模化落地的核心瓶颈，**已从硬件和底层运控转向了数据。**

## 亿欧智库：具身大脑进化严重受阻于数据的结构性困境

- 2026年4月，斯坦福 HAI 发布的《AI Index Report 2026》揭示了具身智能的残酷现状：机器人操控在仿真环境中的成功率高达 89.4%，但在真实家庭场景中骤降至 12%。
- 77个百分点的“仿真到现实迁移鸿沟”（Sim-to-Real Gap），让数十亿资金堆出的“机器人大脑”在真实环境中沦为“行动废柴”。

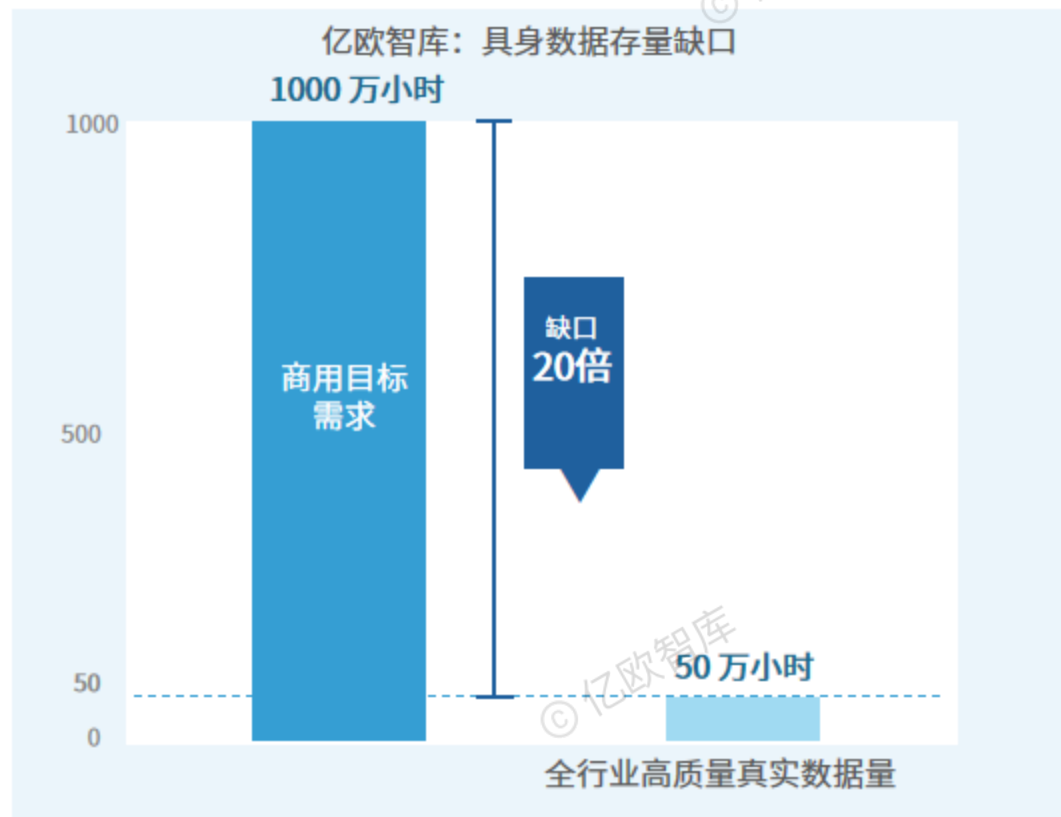
亿欧智库：机器人操控成功率对比



当前人形机器人产业面临“本体与小脑技术趋于成熟，但大脑进化严重受阻于数据”的结构性困境。

## 亿欧智库：具身数据存量严重不足的现状

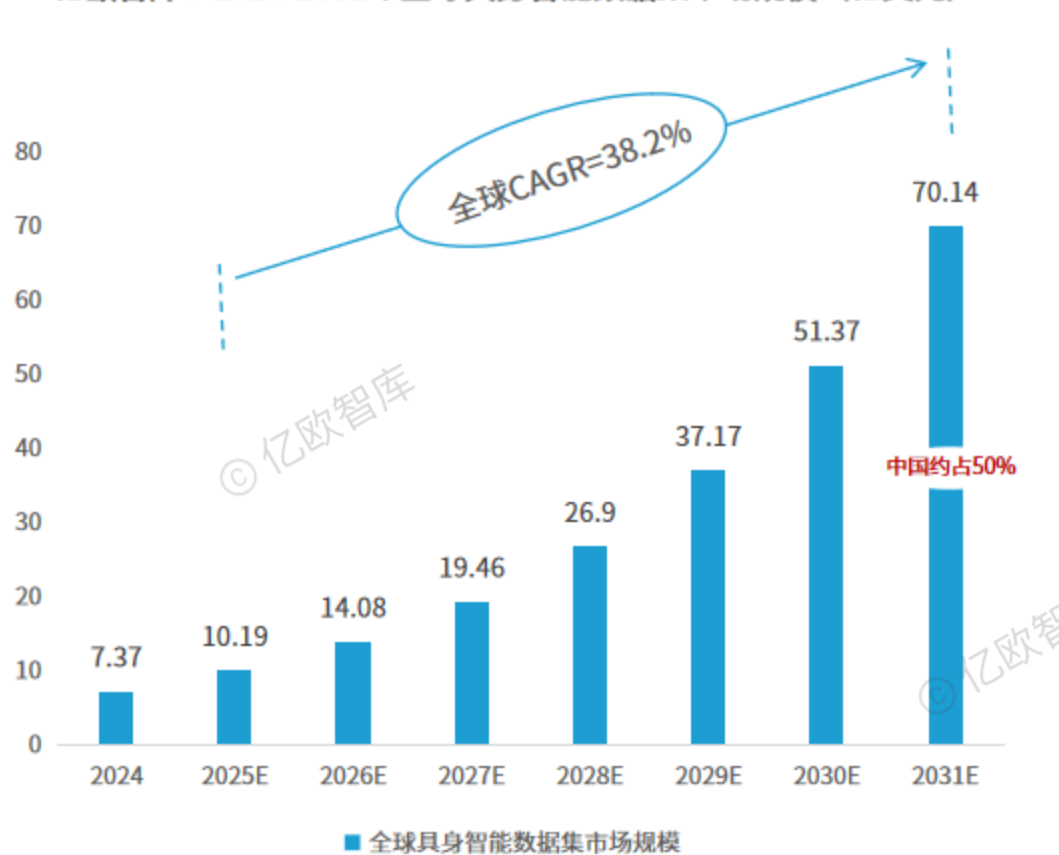
- 具身智能要达到真正可用，至少需要1000万小时级别的真实场景交互数据，然而当前全行业高质量真实物理交互数据总量仅约50万小时，数据缺口高达20倍。



## 1.2 中国具身智能数采与数据市场规模与驱动因素

- ◆ 2024年全球具身智能数据集市场规模大约为7.37亿美元，预计2031年达到70.14亿美元，2025-2031期间年复合增长率（CAGR）为38.2%。中国具身智能数据集市场快速发展，预计到2031年占全球50%左右。
- ◆ 2026年，中国具身智能数据采集行业正在四个维度同步发生结构性变化：行业标准体系从零开始加速构建；数据要素从产业附属品蜕变为独立市场化商品，形成明确的价格信号和交易机制；数据生产从分散手工走向工业化规模产出；资本以前所未有的密度涌入数据基础设施赛道，估值体系快速重构。

亿欧智库：2024-2031年全球具身智能数据集市场规模（亿美元）



### 四大驱动因素

1

#### 标准体系加速构建

国家标准化管理委员会正式下达《高质量数据集 具身智能 面向训练基地的数据采集与模型训练规范》国家标准计划，浦东落地首个国家级标准化试点。

2

#### 数据要素市场化

京东、百度上线数据交易平台，觅蜂科技首发467个数据集，真机数据定价500-1000元/小时，数据从成本项变为独立商品。

3

#### 数据规模工业化

觅蜂目标千万小时级产能，京东发动10万员工采集1000万小时，全国近30家训练场，数据生产从作坊走向工厂。

4

#### 资本极速涌入

Q1融资近300亿元，同比+63%；光轮智能估值20亿美元，它石智航单轮4.55亿美元，资本押注数据基础设施赛道。

## 1.3 中国具身智能数采与数据行业的核心挑战与瓶颈

- ◆ 中国具身智能数据采集行业的核心挑战，本质上是一个环环相扣的“五环困局”：成本高企限制了数据供给的规模，效率低下拉长了产业化的时间窗口，异构壁垒阻碍了数据的流通复用，多模态复杂度推高了技术和人才门槛，而标准化缺失则让上述所有问题难以在系统层面得到解决。如何让成本、效率、异构、复杂度、标准化五个维度从相互制约的“死结”转变为相互促进的“飞轮”，是2026年具身智能“数据元年”需要回答的最核心命题。



## 目录

## CONTENTS

### 01 中国具身智能数采与数据产业发展背景

- 1.1 中国具身智能数据行业发展背景
- 1.2 中国具身智能数采与数据市场规模与驱动因素
- 1.3 中国具身智能数采与数据行业的核心挑战与瓶颈

### 02 具身智能数据采集技术路线对比

- 2.1 遥操作：数据质量最高，成本最大
- 2.2 EGO 第一人称数据：成本较低，规模化潜力最大
- 2.3 便携 UMI：成本与质量的最佳平衡点
- 2.4 仿真数据：产能无限，但 Sim2Real 鸿沟难逾

### 03 具身智能数采与数据行业产业链结构

- 3.1 底层基础设施层
- 3.2 数据采集层
- 3.3 数据加工层
- 3.4 数据应用层
- 3.5 优秀企业案例分析
- 3.6 数据采集案例 TOP20
- 3.7 产业图谱

### 04 具身智能产业商业化突围策略与未来展望

- 4.1 商业模式演进与盈利路径
- 4.2 未来趋势
- 4.3 战略建议

## 2 具身智能数据采集技术路线评价维度

◆本节将具体分析具身智能数据采集技术路线：遥操作、EGO第一人称、便携UMI、仿真数据。

◆在具身智能数据采集中，质量、成本、规模、通用性不可兼得，为提供更好的分析各采集技术，以数据质量、规模化、成本效率、跨机型复用四个维度进行评价。维度具体考量因素如下。

评价维度	定义	核心衡量要素
数据质量	对VLA模型训练的直接价值	动作标签精度（关节角度、位姿等）、多模态同步（视觉、语言、动作、力触觉等）、物理真实性（摩擦、形变、光照变化等）、任务多样性（长尾场景等）、噪声水平（标注错误、时间戳出错位等）等
规模化	单位时间内可采集的数据量及扩展上限	单实例采集速度、并行扩展能力、人力依赖度、时空约束、数据生命周期等
成本效率	单条有效数据的综合成本	硬件、人力、后处理成本、边际成本、失败率成本等
跨机型复用	数据是否可迁移到不同构型的机器人的可用性与迁移损耗	动作空间标签、观测空间一致性、形态差异容忍、策略泛化性等

评估方法



案头研究



企业访谈



专家评估



模型分析

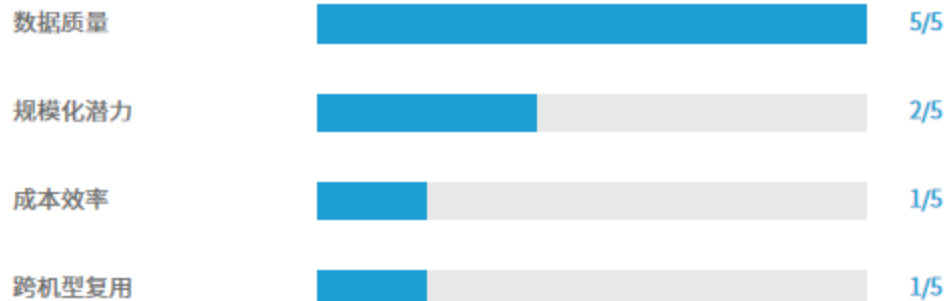
## 2.1 遥操作：数据质量最高，成本最大

◆ 遥操作（Teleoperation）是当前具身智能数据采集中质量最高的方案，行业公认的“黄金标准”。其核心逻辑是由人类远程操控机器人本体完成具体任务，全程同步记录关节轨迹、力反馈、触觉信息和视觉画面，并完整保留人类在复杂环境中的决策过程——包括如何判断、取舍并不断修正动作。然而，遥操作也是成本最高、效率天花板最低的路线，被业界称为具身智能数据生产的“重工业模式”。

亿欧智库：遥操作采集流程及成本结构拆解



数据质量最高，成本最大



### 采集流程

场景准备：布置操作环境，放置目标物体，调整相机角度，确保光照充足。

设备校准：穿戴动捕服、校准传感器，验证关节映射，测试通信延迟。

任务演示：操作员执行任务，系统记录RGB图像、深度图、关节角度、末端位姿、力/力矩等数据，自动标注成功/失败。

**适配场景：**高精度精密装配、复杂长程任务，是模型落地“最后 10%”高质量资料来源

### 成本结构拆解

- 设备购入 > 20 万元\套
- 采集元人力成本约300元\天
- 真机数据市场价500-1000元\小时

### 数据质量评级

关节位姿精度	★★★★★
力触觉反馈	★★★★★
多模态覆盖	★★★★★
跨机型复用	★☆☆☆☆
规模化潜力	★★☆☆☆

### 典型代表

- 智元机器人自建了遥操作数据采集体系，其张江工厂部署200台机器人，由专业数据采集员通过遥操作设备远程遥控机器人完成倒茶水、熨衣服、叠被子等任务。

## 2.2 EGO 第一人称数据：成本较低，规模化潜力最大

- ◆ 第一人称视角（Ego-centric）采集被越来越多观点认为是模型训练的“新主食”。其核心思路是采集者佩戴头戴式或颈挂式设备，在日常工作环境中自然完成任务，不依赖特定机器人本体。这一路线具有四条路线中最高规模化潜力和最低的部署门槛。

亿欧智库：EGO 第一人称数据采集流程及成本结构拆解



成本较低，规模化潜力最大



### 采集流程

基于头戴式RGB第一人称视频流  
结合SLAM轨迹恢复、三维手部姿态估计、逆运动学与动作重定向  
构建统一相对末端执行器动作空间  
全域采用头戴主相机+腕部辅相机双视角，保证手部交互可见性。

**适配场景：**大规模冷启动训练，  
正成为基座模型预训练阶段的“主食”

### 成本结构拆解

- 德马科技发布的OmniEgo方案，相较传统数据采集模式，综合成本降低80%，单小时数据成本降至传统模式的1/5。

### 数据质量

- 第一人称视角的数据虽缺乏精细力反馈，但在模型启动训练中发挥了关键的“冷启动”作用。

### 典型代表

- 德马科技联合麦擎智能正式发布第一人称视角（Ego-centric）具身智能数据采集全管线——OmniEgo，旨在突破数据采集低效、成本高、通用性差等障碍，打通具身智能规模化数据供给瓶颈。

## 2.3 便携UMI：成本与质量的最佳平衡点

- ◆ UMI (Universal Manipulation Interface) 由斯坦福大学与Google团队联合提出，是介于重量级遥操作和轻量级Ego采集之间的“平衡方案”。其核心思路是通过一个手持式夹爪装置（装有GoPro相机和IMU传感器），让人类直接操作夹爪采集数据。与传统遥操作不同，UMI不追求实时高精度远程操控，而是专注于高效生成可泛化的机器人学习数据。

亿欧智库：便携UMI采集流程及成本结构拆解



### 采集流程

设备准备与简单校准  
 操作员手持夹爪操作物体，设备实时记录RGB图像、深度图、设备位姿、夹爪开合度等数据。  
 数据后处理包括轨迹平滑、时间对齐和质量检查，数据采完即可使用。

**适配场景：**中等难度操作的规模化真机数据采集，质效平衡最佳方案



### 成本与质量的最佳平衡点



### 成本结构拆解

- 灵生科技发布的LivUMI方案以万元级成本实现高精度具身数据采集。
- 单条采集成本约0.6-1.2美元，远低于遥操作路线。

### 跨机型复用

- UMI路线的重要优势在于跨机型数据复用能力——不依赖特定机器人本体硬件，一套采集数据可适配多种机器人形态。

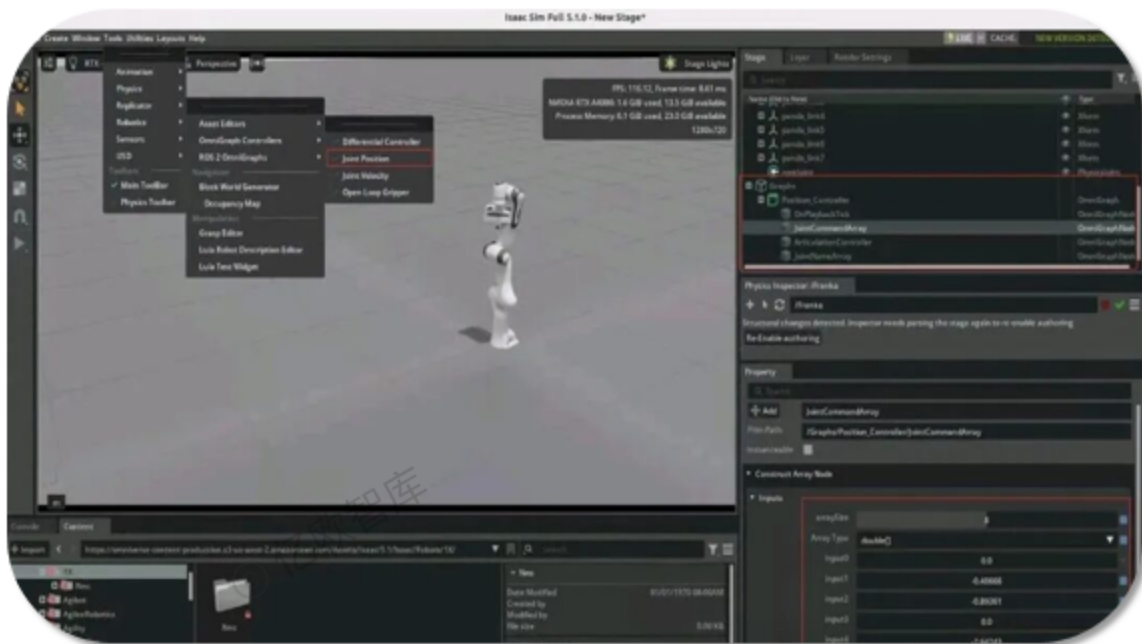
### 典型代表

- 灵生科技发布的LivUMI方案秉持“人本采集”理念，主打“极致性价比、工业级品质、开箱即用、适配主流本体、开放数据空间管线”等核心优势，旨在打造普惠型数据采集方案。

## 2.4 仿真数据：产能无限，但Sim2Real鸿沟难逾

- ◆ 仿真数据是四条路线中产能上限最高、成本最低的方案——理论上只要有足够的算力，就能无限产出标准化训练数据。然而，仿真数据的核心瓶颈并非产能，而是“Sim2Real迁移鸿沟”：虚拟环境永远无法完全模拟真实世界的物理复杂性和各种意外。

亿欧智库：仿真数据采集流程及成本结构拆解



产能无限，但 Sim2Real 鸿沟难逾

数据质量	<div style="width: 25%;"></div>	2/5
规模化潜力	<div style="width: 100%;"></div>	5/5
成本效率	<div style="width: 100%;"></div>	5/5
跨机型复用	<div style="width: 100%;"></div>	5/5

### 细分技术路线

- 基于物理引擎的仿真合成数据生成
- 基于人类操作视频的仿真数据生成
- 仿真与真实数据融合的闭环采集
- 基于生成式模型的仿真数据生成

**适配场景：**海量标准化场景的大规模预训练，必须与真实数据配合使用

### 成本结构拆解

- 仿真数据的边际成本趋近于零，仅需承担算力开销。

### 数据质量

- 仿真→真实的迁移鸿沟高达77个百分点：机器人在仿真环境中操作成功率达89.4%，进入真实场景便断崖式跌至12.4%。

### 典型代表

- 无问智科以真实数据为锚，绕开Sim2Real鸿沟，实现“真机采集→无限仿真泛化→虚实融合产品交付”，实现了仿真数据的物理可信度与规模化之间的平衡。

## 目录

## CONTENTS

### 01 中国具身智能数采与数据产业发展背景

- 1.1 中国具身智能数据行业发展背景
- 1.2 中国具身智能数采与数据市场规模与驱动因素
- 1.3 中国具身智能数采与数据行业的核心挑战与瓶颈

### 02 具身智能数据采集技术路线对比

- 2.1 遥操作：数据质量最高，成本最大
- 2.2 EGO 第一人称数据：成本较低，规模化潜力最大
- 2.3 便携 UMI：成本与质量的最佳平衡点
- 2.4 仿真数据：产能无限，但 Sim2Real 鸿沟难逾

### 03 具身智能数采与数据行业产业链结构

- 3.1 底层基础设施层
- 3.2 数据采集层
- 3.3 数据加工层
- 3.4 数据应用层
- 3.5 优秀企业案例分析
- 3.6 数据采集案例 TOP20
- 3.7 产业图谱

### 04 具身智能产业商业化突围策略与未来展望

- 4.1 商业模式演进与盈利路径
- 4.2 未来趋势
- 4.3 战略建议

## 3.1 底层基础设施层

- ◆ 底层基础设施层是构建高吞吐、低延迟、可拓展的具身数据基座，承接数据从采集、传输、存储到加工训练的全流程物理与系统环境。其作为整个数据产业链的**物理底座**，包含算力、存储、网络和基础软件等。
- ◆ 具身数据面临三大核心挑战：**数据量极大**、**实时性要求极高**、**多模态强异构性**，这些挑战对采集端写入带宽、训练集群互联带宽、端到端操作延迟、数据加载访问延迟等设计了硬性指标。

亿欧智库：具身智能数采与数据行业—底层基础设施层



### 基础设施的挑战与对策

- **数据洪峰**：单台机器人1小时数据达1-3TB。对策：边缘缓存+异步上传+并行文件系统。
- **实时性要求**：遥操作延迟超过50毫秒就会失败。对策：本地闭环控制+QUIC/UDP定制传输。
- **多元数据的时间对齐**：多种传感器时间不同步。对策：硬件PTP同步+软件时间戳归一化服务。

### 关键KPI

- 采集端写入带宽：支持高吞吐写入（高端系统目标 $\geq 3\text{GB/s}$ ）
- 节点内GPU互联带宽  $\geq 200\text{GB/s}$
- 消费级/VR遥操作端到端延迟  $\leq 50\text{ms}$
- 数据加载访问延迟  $\leq 100\text{ms}$ （综合）

注：所列KPI基于行业头部企业公开与主流技术方案整理

## 3.2 数据采集层

- ◆ 数据采集层承担从物理世界到数据世界的**翻译功能**。将人类真实场景操作数据转化为高质量、多模态的原始数据，提供具身智能所需数据信息。
- ◆ 数据采集层包含**采集团队**、**硬件设备**以及**采集平台**，三者属于协同关系，其中，采集团队保证采集的质量、效率和可重复性；硬件设备提供高保真、多模态、高精度实时同步的物理感知能力；采集平台统一调度、控制、记录、监控所有采集任务。最终产出带完整元数据的原始多模态数据包。

亿欧智库：具身智能数采与数据行业-数据采集层

### 采集团队

包括遥操作专家、动捕演员、任务设计师、质量审核员等

核心：**保证数据质量**，如专家提供黄金标准示范，审核员实时剔除坏数据



### 采集平台

负责低延迟遥操作传输、多模态数据同步录制、实时监控仪表盘，以及标定工具和数据上传管道。

### 硬件设备

机器人本体，视觉、力觉、触觉传感器以及边缘计算节点。

关键指标：传感器同步误差小于1毫秒，写入带宽超过3GB/s

### 采集团队

- 定义采集要完成的动作或任务清单
- 操作员、专家等进行执行或记录
- 质量审核员实时进行质量审核
- 轻量级预标注

### 硬件设备

- 先对所有传感器进行参数校准
- 所有传感器同时工作
- 边缓边存
- 对原始数据进行无损或有损压缩

### 采集平台

- 通过采集软件平台，将设计好的任务列表推送到各个采集终端
- 采集各数据流并分文件记录
- 提供实时可视化界面
- 上传中央存储集群

### 3.3 数据加工层

- ◆ 数据加工层的核心是将采集来的原始、多模态、非结构化数据转化为**高质量、带标注、对齐好的**训练就绪数据集。主要分为清洗、标注、治理、管理平台等。
- ◆ 具身数据采集成本极高，可通过数据增强低成本扩充数据集，提升模型泛化性，常见的有视觉增强，物理参数随机化、运动增强和场景增强，增强后的数据保留原始标签并标记。加工后的数据进入模型训练前，需要生成数据卡片，产出直接用于训练、可复现、可追溯的标准数据集。

亿欧智库：具身智能数采与数据行业-数据加工层



#### 数据增强

通常在标注后、治理前，具身数据增强必须保持物理一致性

- 扩充数据集规模，降低采集成本
- 提升模型泛化性，让模型适应不同光照、纹理、物理参数等变化
- 弥补采集偏差，例如训练集中只有“红色方块”，通过颜色增强得到“蓝色方块”
- 增强后的数据必须继承原始标签，并标记

#### 数据集构建/版本管理

将加工后的数据组织成规范化、可复现、可追溯的数据资产

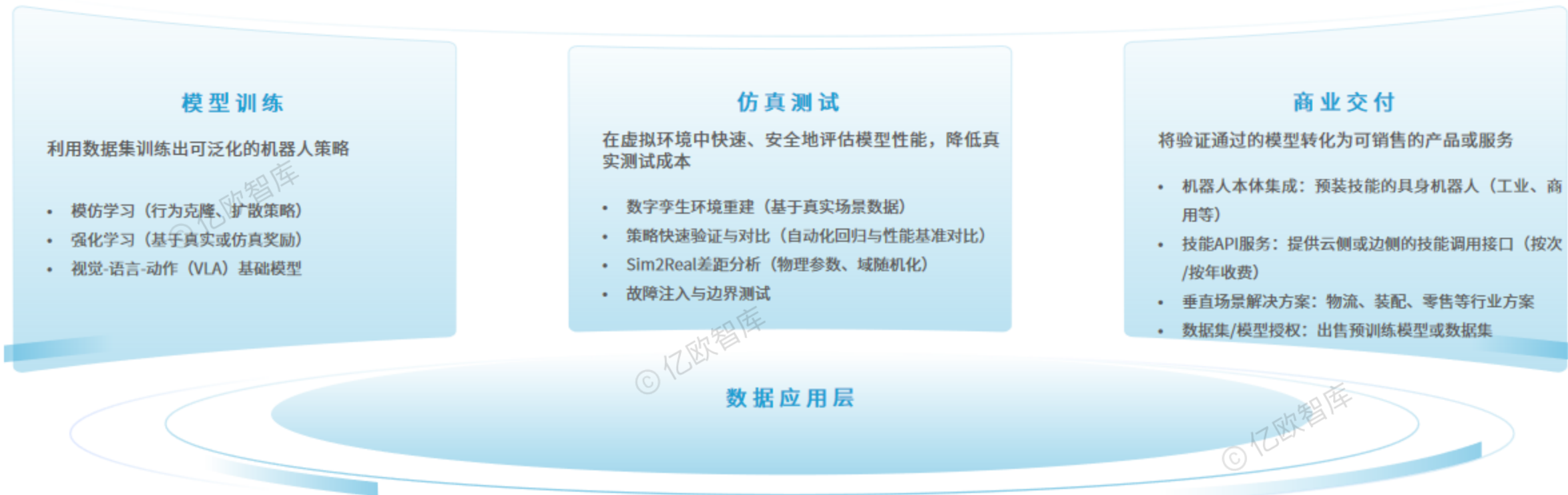
- 标准化输出：让数据格式、元数据、目录结构统一
- 版本控制：记录每一次数据变更，支持回滚和对比
- 可复现性：保证同一数据集版本在不同时间、不同机器上产生相同的训练效果。
- 数据卡片 (Data Sheet)：提供详细文档，包括来源、传感器、任务分布、标注规则、伦理声明等。

管理平台：贯穿整个流程，将以上流程自动化、可视化。负责流水线编排、标注工作台、数据版本管理与血缘追踪、进度监控、自动化集成等

### 3.4 数据应用层

- ◆ 数据应用层的核心是将加工好的高质量数据集，转化为可部署的智能能力，最终实现商业化。关键环节包含**模型训练**、**仿真测试**、**商业交付**，三者形成以数据为纽带的迭代闭环。
- ◆ 模型训练作为核心研发环节，主要利用标准化、可追溯的高质量数据集开展策略/技能训练，仿真测试既作为合成数据生成器扩充训练集，又作为安全测试场快速评估模型泛化性与鲁棒性，最终将训练好的模型以产品/服务形式交付给客户实现价值变现，同时现场产生的真实有效数据回流至训练端，驱动模型持续迭代。

亿欧智库：具身智能数采与数据行业-数据应用层



### 3.5.1 无问智科——世界模型驱动的物理AI数据基座定义者与打造者

- ◆ 浙江无问智行科技有限公司（下称无问智科）成立于2022年，始终坚守“关注真需求、解决真问题”的核心理念，提供支撑场景数据采集、数据处理、模型训练及评测的全链路闭环基础设施能力，构建面向物理AI的数据闭环。帮助客户解决“用什么数据、怎么处理数据、如何验证数据价值”的全链路问题。
- ◆ 作为国内领先的物理AI数据基础设施企业，业界首个世界模型驱动的物理AI数据基座，无问智科围绕具身智能、机器人与自动驾驶领域，构建了覆盖“采集—生成—仿真—测评与验证”的全流程闭环体系，并推出业内首个物理AI数据基座平台“无垠”，全栈自研“无垠”物理AI数据基座平台，精准破解行业“数据、场景、工具链、测评标准”四大痛点，构建起行业领先的技术体系。

亿欧智库：无问智科“无垠”物理AI数据基座平台



### 3.5.1 无问智科——构建从数据采集、处理、训练、仿真到验证的完整闭环

- ◆ 无问智科建立了实体数据采集训练场，该训练场拥有8000平方米固定场景、505亩室外场地与937平方公里全域开放场景，日产数据上千小时，合成数据万级规模化，仿真+真机测评验证上万次。以该训练场为平台，深度链接数十家真实终端场景，形成高价值、规模化的场景生态矩阵，全面覆盖物流仓储、家庭服务、酒店文旅、工业制造、办公服务、商业零售6大核心场景。
- ◆ 具身智能数据采集训练场以“虚实融合闭环”、“多层次全域场景开放”、“市场化全链共建”三大特色为“无垠”数据基座平台提供了坚实物理支撑，形成了一条从“数据采集”到“训练”再到“产业赋能”的完整生态链。

#### 亿欧智库：无问智科核心数据闭环体系技术壁垒

##### AI原生数据处理管线



领先的多模态世界模型理解与高精度场景重建（毫米级精度），实现高效自动化标注与时空标定（1-2像素误差）

##### 虚实结合场景生成能力



业内首个交付生成式仿真产品，已落地头部车企，覆盖真机数据处理+仿真场景泛化搭建，具备虚实联动优势

##### 深度评测认知体系



拥有200+评测指标，能指导客户基于评测结果定向采集数据以迭代模型，具备反向数据规划指导能力

##### 算法与工程结合能力



具备将复杂算法可靠落地为稳定软件工具的工程壁垒，确保算法在实时性、可靠性上满足工业级要求

#### 无问智科核心数据闭环体系支持私有化交付

- 多源异构数据采集管理平台：面向全部客户提供私有化部署，支撑规模化、标准化的数据采集、数据处理到模型验证的全链路；
- 跨本体通用数据迁移工具链：主要服务战略客户，实现多范式、跨本体数据互通转换；
- 自动化数据处理管线：覆盖采集-清洗-切边-标注全流程，整体周期压缩至小时级别；
- Real2sim高效场景构建管线：场景重建资产搭建效率提升3-4倍，耗时仅为原来20%-30%；
- 规模化任务虚实融合评测能力：基于规模化评测体系，支撑具身大脑模型优化和迭代路径，设计和定义最优场景和数据方案，指导客户“该补什么数据、怎么补”。

无问全栈自研数据管线效率持续迭代，整体数据处理效率每三个月提升近100%，VLA/WAM标注时间缩短至原来的1/10，数据合成效率缩短至原来1/100。

#### 亿欧智库：无问智科具身智能数采训练场六大核心作业场景

##### 物流仓储



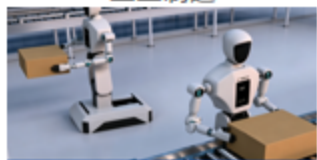
##### 家庭服务



##### 酒店文旅



##### 工业制造



##### 办公服务



##### 商业零售



#### 亿欧智库：无问智科真实商业运营场景数据采集案例

收纳整理



物流分拣



生产制造



物品配送



构建数据采集+处理+标注一站式数据解决方案：

- 数据类型：多模态数据采集覆盖视觉、语言、触觉、嗅觉等多种感知数据
  - 数据处理：多模态数据清洗、时空对齐、数据重建与拼接、数据质检等全自动化
  - 数据标注：从采集规范完成的Task(Action)加速标注理解VLA语义，提升标注准确性和效率，自研标注平台可快速适配客户需求
- 有效数据采集率稳定达80%以上，不同场景采集300-1000条/天/台

## 3.5.2 数据堂——专业具身智能数据采集和数据标注服务商

- ◆ 数据堂成立于2010年，早期专注于传统机器学习领域的小数据服务，通过不断拓展和发展，24年正式布局具身智能数据，聚焦大模型与具身智能的交叉学科数据，25年全面进入具身智能本体数据采集与标注领域。公司依托**柔性制造生产线技术壁垒**，提供**版权数据集、定制服务及软件解决方案**三类产品，服务于通用大模型公司、具身大脑算法公司、机器人本体公司以及垂直应用公司。
- ◆ 数据堂具身智能数据采集通过“**场内模拟+场外真实**”的双轨采集策略，覆盖**零售、家居服务、仓储、医疗、工业**等多类真实场景数据。

### 亿欧智库：数据堂具身智能数据业务模式

#### 版权数据集



**环境数据**  
2.88亿组3D模型和场景数据，涵盖人物、动物、植物、建筑物和日常物品。



**决策数据**  
10万组人物第一人称互动视频标注数据，涵盖烹饪、手工、运动等。



**控制数据**  
15万段具身智能灵巧手数据，涵盖视觉（RGB-D）、关节状态及力/力矩反馈。

#### 定制化采集与标注



**多形态**



**多场景**

**多类型**

- 家庭看护数据
- 工业运输数据
- 餐厅送餐数据
- 药店/超市物料上架数据
- 超市导航数据
- ...

针对客户特定硬件及场景需求，利用全球实施团队及众包资源进行定制化采集。例如，针对不同本体的跨本体数据需求，提供专属的定制化服务。

#### 软件解决方案



**感知标注**



**VLA标注**

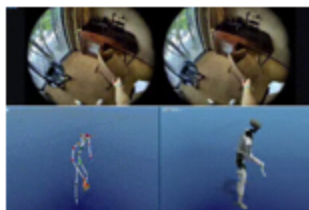
提供经过大规模复杂任务验证的数据标注平台、管理平台及治理工具，支持域内处理

### 亿欧智库：数据堂具身智能数据采集

#### 场内模拟采集 + 场外真实采集

扩展采集工厂至8000平方米，具备零售、家居服务、仓储、医疗、工业等多类真实场景的模拟环境，并部署300套灵巧手操作设备及多形态机器人本体。覆盖抓取、放置、搬运、装配到分拣、递送等数十类操作任务。

采集设备	真实场景	任务类型	多模态传感器
灵巧手、遥操作机械臂、力反馈交互设备、惯性动捕系统、多视角RGB-D视觉设备等	零售、家居服务、仓储、医疗、工业等多类型环境	抓取、放置、搬运、推拉、分拣、整理、装配、递送、交互等	第一视角视频数据、动作轨迹、关节角度与力反馈信号



- 在入户或在真实商业场景中进行，强调“真实的人”在“真实的场景”中操作。要求操作人员具备相应职业背景（如专业保洁员），且单一场景采集时间不超过8小时以保证场景多样性。
- 超1000台专业穿戴采集设备，实现单台设备日产出3.5小时有效数据，规模化成本控制在100元/小时。覆盖烹饪、手工、清洁收纳、运动等室内外百余种真实操作场景，还支持从纯视频到双目+点云+关节的多模态采集方案。

### 亿欧智库：数据堂Ego-Centric数据项目案例



- 项目概述：客户购买部分“10万小时多场景Ego-Centric数据”成品数据后，提出多场景定制采集方向，覆盖家居、办公等多个场景的真实数据需求，其中家居室内场景是本次项目的重点之一。
  - 围绕家居室内场景的定制需求，数据堂根据项目标准租用了多个不同布局的真实住宅空间，用于Ego-Centric数据采集，覆盖厨房、客厅、卧室、书房、卫生间、阳台等典型家庭空间。
  - 通过多户型、多空间、多任务的采集设计，建立从场地协调、任务拆解、人员培训、设备管理、采集执行、质量检查和数据交付的完整流程，保证不同住宅空间、不同采集人员、不同任务类型统一的操作标准和质检要求。
- ✓ 项目成果：周产能有效数据约5000小时

### 3.5.3 公象智能——从数据定义到API交付的一体化数据交易平台

- ◆ 公象智能作为慧用工的全资子公司，继承了慧用工“基层就业人才池+全国物理采集网络+自研SaaS+企业服务经验”4大关键资产，并投向具身智能数据赛道。
- ◆ 公象智能致力于将数据生产仓“人力驱动”升级为“系统驱动”，搭建了一体化数据基础设施平台，建立了从数据定义→采集→处理→标准化→API服务的全链路系统。公象数据交易平台具备标准化生产、可下单可比价、可复用可变现、规模化交付等特点。

亿欧智库：公象智能具备的4大关键资产



- 4大关键资产
- 01 基层就业者人才池  
300万+，覆盖外包/派遣/招聘/灵活用工全业务
  - 02 全国物理采集网络  
300城，近20分公司，任意城市24小时内组队
  - 03 数智化运营系统  
自研SaaS，10年SaaS技术沉淀，任务分发/质检/结算全自动
  - 04 企业服务经验  
1.2万+，含美团等大客户，合作资质完善

数据来源：公象、亿欧智库

亿欧智库：公象智能一体化数据基础设施平台



从数据定义→采集→处理→标准化→API服务的全链路系统  
将数据生产从“人力驱动”到“系统驱动”

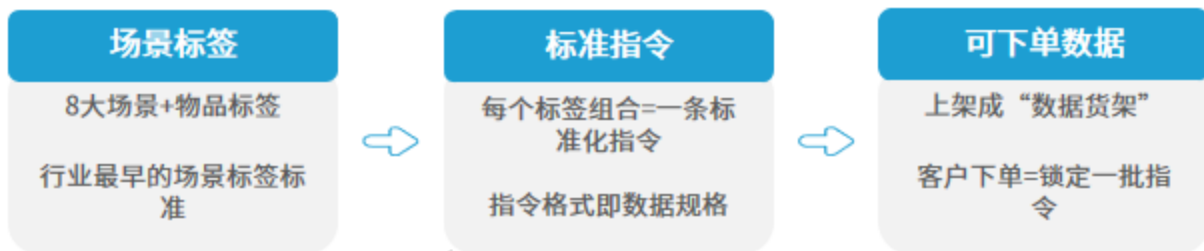
亿欧智库：公象智能数据交易平台特点



### 3.5.3 公象智能——具身智能数据交易平台

- ◆ 公象数据交易平台核心理念是**定义数据**，将物理世界的“人、物、空间”映射为可计算、可交易的数字化数据商品。公象将场景及内部物品标签建立标签体系，每个标签组合对应一条标准化指令，每条指令=一条数据规格=一件可下单的商品，其指令库覆盖**8大重点场景**，**600+**细分场景，**百万+**标准任务，**上亿条**可下单指令。数据交易平台把指令库变成数据货架，形成浏览、下单、取货、计费的一站式**数据交易**。让数据采集像电商订单一样可下单、可调度、可交付。
- ◆ 公象智能在定义数据的同时，将任务中的复杂场景拆解为可复制、可派单、可验收的标准任务单元，实现跨场景持续复用，即一套体系支撑多场景持续扩展，每条指令都能产生百万级数据。同时通过专职采集以及即将依托慧用工10年人力基建形成的众包网络，形成采集网络，实现从可控生产到规模化扩展，让采集从依赖“人”到依赖“系统”。

亿欧智库：公象智能数据交易平台核心理念及核心能力



客户用“指令”预定数据，指令库本身就是数据库的金矿——数据提前产生、可多次复用  
**8大重点场景 X 600+细分场景 X 百万+标准任务 = 上亿条**可下单指令

**01 数据货架 标准数据集上架**  
按8大场景/模态、规模分类上架，即选即购

- 多模态case数据集
- 场景标签检索
- 规模/价格透明

**02 指令预定 按需定制采集**  
指令=一条数据规格，下单=锁定一批指令。  
系统排产采集

- 百万级标准指令库
- 上亿条可下单指令
- 客户下单即提前产生数据

**03 多形态交付 数据即服务**  
交付形态：**标准数据集/API数据服务/定制采集**

- 数据集打包下载
- API订阅调用
- 私有化定制采集

**04 交付与计费 安全可信交易**  
交付方式：**现货直购/预定采集/月度框架，统一按小时计费**

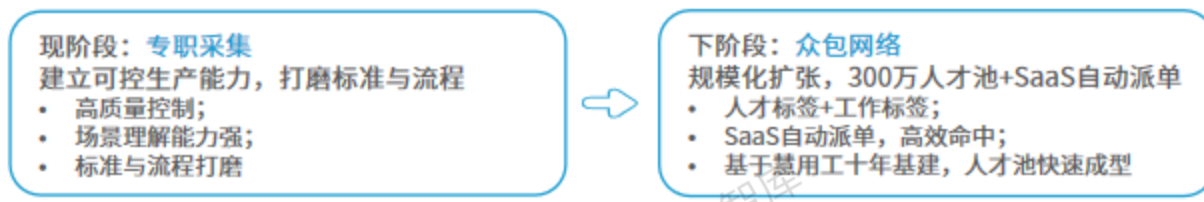
- 统一按小时计费
- 数据版权与脱敏合规
- 数据可复用增值

亿欧智库：公象智能8大重点采集场景



标签=数据标准 任务系统=可复制能力  
 从标签到数据：自研场景标签体系 - 标签泛化生成指令模板 - 任务标准化拆解 - 跨场景持续复用

亿欧智库：公象智能采集网络



- 公象同时建立了**指令标准**和**人才标准**，指令侧将场景拆解为可下单的指令；人才侧把人才匹配变成系统作业，两套标准把不可控的代采，变成了可批量下单的标准化生产，标准化、规模化的采集作业，让采集不再依赖“人”，而依赖“系统”。
- 在数据采集的源头端，公象实现行业**两倍以上**的有效数据产出效率，10个小时的采集时长其有效数据为3-4小时，有效率约35%，是行业平均的2倍以上。

### 3.5.4 觅蜂科技——一站式物理AI数据服务平台

- ◆ 觅蜂科技是智元机器人孵化的一站式物理AI数据服务平台，致力于打造具身智能数据的平台型供给基础设施，实现数据的体系化、标准化、规模化供给，并通过发起蜂巢数据共创行动，联合全球伙伴打造高效的数据流通网络，加速具身智能价值落地。
- ◆ 觅蜂科技不止提供数据，也致力于构建物理AI数据基础设施，实现真机遥操、无本体采集、仿真数据全范式覆盖，并打通硬件、软件、平台、运营全链路，以类型全、质量好、交付快的数据供给能力，让高质量数据像水电一样即取即用。

#### 亿欧智库：觅蜂科技核心优势

##### 全类型数据覆盖

真机遥操  
无本体采集  
仿真数据

##### 全场景产能规模

涵盖工厂、物流、超市、家庭等  
全域场景  
预计2026年实现千万小时级产能

##### 工业级质量保障

全流程闭环运营：任务设计  
→标准化采集→人工在环审核  
→算法验证→标准交付

##### 全栈自研能力

硬件+软件+平台（数据商城）  
+全球采集网络闭环

##### 顶尖学术认可

开源数据集 Hugging Face 累计下载  
110w+, Github Star 2.9k+, 被 MIT、  
清华、伯克利、哈佛等采用

##### 蜂巢数据共创

联合上电科、国家数据标委  
会等发起，打造开放高效的  
物理AI数据流通网络。

#### 亿欧智库：觅蜂科技蜂巢数据共创行动

产业链伙伴：数据采集服务商、硬件厂商、场景方、AI模型企业  
科研与机构：科研机构、高校实验室、产业联盟、行业协会

##### 共享资源

接入全球供需网络  
共享海量订单

##### 共建标准

统一数据标准  
降低写作成本

##### 生态合作

优先参与软硬件  
与数据集生态建设

##### 行业发声

联合发布白皮书  
共建话语权

##### 产业推动

共享技术自办  
推动产业爆发

#### 亿欧智库：觅蜂科技无本体数据采集与治理基础设施

##### MEgo Gripper



无线UMI标杆产品，  
支持全场景高精度物  
理交互采集，实现高  
保真、高精度、全场  
景数据采集。

##### MEgo View



多视角全感官数据采集  
终端，融合头部300°全  
景与腕部细节捕捉，实  
现高精度、普惠化数据  
生产。

##### MEgo Engine



一站式数据治理服务  
平台，实现数据预处  
理、空间感知、质量  
评估、数据标注。

##### G2 Air



作业与数采一体化，  
人机合一，在完成任  
务时采集高质量数据，  
可在宽度800mm通道  
完成原地掉头。

MEgo x G2 Air 原生同构，高效部署：无本体采集数据和真机数据同源共生，为模型训练提供高质量、无差异的数据样本。基于MEgo系列的同构型传感器和夹爪一致性数据训练模型，可无缝对接部署到G2 Air，快速实现机器人的自主作业能力。

#### 亿欧智库：觅蜂科技仿真数据



物体资产 环境资产

- 觅蜂科技拥有8000+物体资产与仿真场景，实现用户侧场景重建、专家轨迹数据获取、评测结果可视化。包含9大类材质、180+类产品、12类原子技能。
- 觅蜂科技通过数字孪生级仿真环境+自然语言驱动场景生成+智能仿真数采，实现高效规模化扩量，仿真与真机评测差异控制在10%以内，应用于多个真实机器人作业环境。

### 3.5.5 宇树机器人——硬件 + 数据流水线 + 模型训练 + 开源数据集的全栈解决方案

- ◆ 宇树机器人在具身智能数据采集（数采）已形成硬件 + 数据流水线 + 模型训练 + 开源数据集的全栈解决方案。其解决方案由人形机器人本体、数据采集共聚、模型训练及推理工具构成。
- ◆ 除此外，宇树全量开源完整的全身遥操作真机数据集以及预训练模型，该数据集完整记录了人形机器人从双足行走、重心平衡动态调整到指尖精细操作的全身协同行为流。截至2026年3月，该数据集合计189万条动作轨迹数据，覆盖衣物收纳、家电操作、餐具整理等多个日常家居与工业生产场景，后续也将持续高频滚动更新。

#### 亿欧智库：宇树机器人数采训练全栈解决方案

人形机器人本体



##### 更高自由度的本体

整机自由度(不含末端): 19  
手臂自由度: 7×2  
腰部自由度: 2  
立柱自由度: 1  
底盘自由度: 2

##### 更大范围的作业空间

采用轮式与升降相结合的移动升降设计  
垂直作业空间: 0-2m  
腰关节运动空间:  
Z±155°、Y-2.5°~+135°

##### 更低延时的控制响应

升降精度: ±0.5mm  
末端夹爪精度±0.1mm  
\*不同末端配置精度有差异  
系统遥操延时: 100ms  
采集频率: 60Hz

数采流程: 新建采集任务-任务编辑与分配-数据采集与标注-数据上传与审核-数据存储-数据导出

数据采集工具

##### 可视化模板管理采集高效

通过模板化配置，一键生成数据采集任务，全流程状态实时跟踪

##### 多样化本体及采集末端，灵活配置

多样化兼容，支持多种机器人与末端配置的数据采集，打通多元设备到高质量模型训练数据。

##### 高并发架构，规模可拓展

支撑数百台机器人同步开展数据采集。确保海量数据流的实时接收与处理。

##### 7X24小时在线采集，稳定可靠

7×24小时可靠采集，具备格式兼容能力，采集数据可直接输出或转换为各类主流训练格式。

模型训练及推理工具

##### 生态丰富，支持主流模型

开放式的模型生态，内置社区数据集，支持开源数据集训练，深度整合PI、GROOT等开源模型。

##### 仿真环境，模型评测

内置高保真、高精度的3D资产库，通过构建仿真环境，快速生成覆盖全面的评测方案，为算法验证提供支撑。

##### 简单易用，快速部署

开箱即用，“一键训练”快速启动模型开发，并利用集成的仿真工具进行可靠评估，最终实现算法到真机的无缝迁移。

##### 分布式训练，高效性能

基于高性能分布式训练架构，实现计算任务的弹性调度与并行加速，按负载资源动态扩展，实现90% GPU利用率。

训练及推理流程: 模型架构选择-训练配置-实时监控-参数编辑-仿真测试-模型导出与部署

UnifoLM-WMA-0: 跨机器人本体的开源世界模型。

决策模式: 分析当前环境状态与任务目标，精准预测机器人与环境未来的物理交互信息。该信息直接辅助策略模块生成动作

仿真模式: 作为交互式仿真器运行，基于机器人动作输入生成高保真的环境反馈，生成高质量合成数据。

#### 亿欧智库：宇树机器人具身智能数采应用场景

服务



生活



商业



工业



#### 亿欧智库：宇树机器人开源具身智能数据集和模型

##### 核心数据集:

G1灵巧手操作数据集、G1夹爪操作数据集、Z1双臂操作数据集；覆盖“单指/多指抓取、双臂协同”等高频工业场景，包含力控、视觉、关节角度等多模态数据。

##### 预训练模型:

主流模仿学习 (IL) 模型: 如DP (Data-efficient Policy)、ACT (Action Chunking with Transformers)，已适配宇树硬件，可直接用于“从人类演示中学习操作技能”。



3月，开源189万条动作轨迹数据，覆盖衣物收纳、家电操作、餐具整理等多个日常家居与工业生产场景，所有数据均来自宇树G1机器人真机采集，且将持续高频滚动更新。

### 3.5.6 银河通用——合成为主、真实为辅，用“仿真世界”喂养“通用大脑”

- ◆ 北京银河通用机器人股份有限公司是市场领先的具身多模态大模型通用机器人企业。其在具身智能数据采集产业选择以合成仿真数据为主、真机数据为辅，独创一套虚实融合训练范式。比如，人形机器人的足式行走、跳舞乃至各类复杂身体控制，都是通过仿真器习得，真实数据的作用仅是提取人类行为形态并在真实世界部署前帮助微调。发布的1.6B 参数的跨本体模型LDA-1B“吃数据不挑食”——仿真的、真实的、人类视频、甚至带噪声的失败数据，都能混在一起有效利用；数据从5000小时扩到30000小时。
- ◆ 目前，银河通用90%的训练数据为虚拟仿真，仅需一周即可生成十亿级操作数据集，成本仅为真实数据的1/100。企业自主研发的VLA（即视觉-运动-动作一体化生成）具身智能大模型，构建起百亿级具身数据集。独特的合成数据训练方式让机器人“手眼脑”更协调，移动、抓取、放置等看似简单、实则高频刚需的任务完成度更高，打造实实在在的“干活派”机器人。

亿欧智库：银河通用“数据金字塔”



银河通用用多本体合成仿真数据做核心，互联网数据和人类行为数据做预训练，只需要极少的真机遥操数据做后训练，就能得到一个高效的机器人具身模型。

- 银河通用在苏州高新区具身智能数据采集中心配备约100台机器人，构建约20个高度仿真的模拟场景，实现年采集千万条级高质量多模态数据。90%的训练数据为虚拟仿真，仅需一周即可生成十亿级操作数据集，成本仅为真实数据的1/100。
- 基于该数据集训练的模型成功率达到99%，远超全球同类具身智能企业，同时实现机器人训练效率比特斯拉高1000倍。

资料来源：银河通用、亿欧智库

亿欧智库：银河通用行业应用及商业化



#### 医疗康养

- 全天候待命，随时执行药品运送及库存管理任务，并为医护人员提供辅助诊断服务。
- 与多家知名三甲医院开展深度合作，推动人形机器人在病房辅助、药房管理与智慧导诊等场景落地。



#### 工业制造

- 在工业产线中，它具备对零部件进行自动分拣、搬运物料、组装、上下料等能力。
- 与宁德时代、博世集团、丰田汽车等国内外领军企业深度合作，累计订单达数千台，帮助企业降低人力成本约30%。



#### 商业零售

- 在零售商超领域，它能迅速而准确地完成取货、交付，并持续进行高效的货物盘点、补货工作，实现7X24小时自主运营。
- 已在全国数十个核心商圈与地标常态化运行，落地20+城、超100家。



#### 家庭服务

- 能够出色地完成清洁打扫、有序整理、准确寻找物品以及拿取快递等工作。

## 3.6 数据采集案例TOP20

序号	企业名称	案例
1	阿米奥机器人	某汽车零部件焊接场景，复杂工艺长期依赖十年以上老师傅。阿米奥机器人到现场后，老师傅演示一遍，机器人便学会，从此长期稳定输出。
2	艾欧科技	艾欧 TeleXperience 已完成乐聚机器人全量遥操作适配，支持全身动作映射与沉浸式数据采集。
3	宸境科技	自研空间智能视觉模组，前置算力提供高精度vslam与AI感知模型，为家庭、工业场景具身智能提供高质量基础视觉感知数据。
4	光轮智能	光轮智能通过工业级规模化仿真评测，将机器人算法验证从依赖真机测试，升级为可复现、可规模化、可持续迭代的研发流程，支持千台机器人并行仿真、单日完成数十万次测试，加速算法优化与真实场景落地。
5	海天瑞声	与国内头部具身人形厂商签署整机数据采集合约，搭建家庭/商用实景采集基地，采集人形机器人导航、人机交互、物体操作多模态连续数据，同步提供仿真数据增强服务。
6	京东	宿迁全国首个具身数据采集社区，发动数十万居民参与，让居民在日常家务劳动的过程中，同步完成数据采集。
7	景联文科技	自建采集基地，复刻数十类不同结构的真实家居场景，配置超3000种不同材质、不同尺寸、不同操作方式的真实家居物品——普通的碗筷、水杯，易碎玻璃器皿、玩具、果蔬等，覆盖了真实家居环境中几乎所有常见物品。
8	灵生科技	LivUMI万元级高效数采方案助力具身智能模型 GEN-1 任务成功率99%、执行速度较行业高3倍。
9	灵初智能	灵初智能推出的 Psi-SynEngine 只用人类数据+2天真机实操可以对标上万小时的真机预训练，轻松学会叠纸盒细致活。
10	里工实业	全新发布ChiroSync如意数采套件，将分布式多模态数采嵌入工业真实作业场景，采集视觉、触觉位姿等多模态数据，并通过“意手乾坤”平台完成数据对齐和后处理，支撑模型训练和真机部署。
11	觅蜂科技	MEgo 无本体采集实现随行即采、全场景覆盖、规模化扩张。MEgo Engine 治理平台，完成多模态数据亚毫秒级时空对齐、清洗、标注、质检。MEgo Gripper 实现1毫米轨迹还原，MEgo View 实现300°全景+腕部特写，并支持亚毫秒级全局同步。
12	穹彻智能	数采硬件落地惠州异构数采厂，覆盖零售、工业、药房、家庭等多场景，形成“无本体数采+遥操作数采”的完整数据生成能力。
13	数据堂	自建8000平具身智能数据采集工厂，覆盖零售、家居、工业、制造等场景。部署300套灵巧手设备及多形态机器人本体，超1000台专业穿戴采集设备，可支撑场外外包模式的大规模数据采集。日产3.5小时有效数据，规模化成本低。
14	无间智科	长三角规模最大、最专业、最领先的具身智能数据采集训练场，“虚实融合闭环”、千万级产能、多层次真实作业场景开放、市场化全链共建。
15	星海图	EDP具身智能开发平台，实现了真实场景下从数据采集、标注到模型部署的全流程闭环。在搬运移动、抓拿放置、封装打包、织物叠放和设备串联已完成千台级订单。
16	奕人科技	基于自动驾驶积累的资源快速切换具身智能领域，与国内头部的物理AI公司有一些合作，26年Q1具身收入已过亿。
17	银河通用	与政府合作建设西南首个机器人数采中心，初步形成“数据采集—模型训练—场景落地”的完整闭环。已组建近百人的专业团队，部署超70台数据采集机器人，每24小时即可生成数万组抓取与操作数据。
18	箬境智能	箬境智能办公、商超、餐饮、家政四大场景具身智能数据集，顺利在江苏省数据交易所完成上架及场内交易。
19	智域基石	帮助传统机器人厂商盘活存量产线数据，新增数据服务收入；改造后数据集可直接用于通用工业操作大模型预训练。将抓取数据有效率从不足10%提升至85%，人力成本下降80%。
20	智元机器人	搭建家居（客厅/厨房）、餐饮、商超、办公全实景复刻场景；百台人形/轮式机械臂同步遥操作采集，日产数万条抓取、打包、清洁、收纳交互数据，累计采集超百万条具身样本。

注：以拼音首字母顺序排序  
资料来源：各公司、亿欧智库

# 3.7 产业图谱



注：部分展示

## 目录

## CONTENTS

### 01 中国具身智能数采与数据产业发展背景

- 1.1 中国具身智能数据行业发展背景
- 1.2 中国具身智能数采与数据市场规模与驱动因素
- 1.3 中国具身智能数采与数据行业的核心挑战与瓶颈

### 02 具身智能数据采集技术路线对比

- 2.1 遥操作：数据质量最高，成本最大
- 2.2 EGO 第一人称数据：成本较低，规模化潜力最大
- 2.3 便携 UMI：成本与质量的最佳平衡点
- 2.4 仿真数据：产能无限，但 Sim2Real 鸿沟难逾

### 03 具身智能数采与数据行业产业链结构

- 3.1 底层基础设施层
- 3.2 数据采集层
- 3.3 数据加工层
- 3.4 数据应用层
- 3.5 优秀企业案例分析
- 3.6 数据采集案例 TOP20
- 3.7 产业图谱

### 04 具身智能产业商业化突围策略与未来展望

- 4.1 商业模式演进与盈利路径
- 4.2 未来趋势
- 4.3 战略建议

## 4.1 商业模式演进与盈利路径

- ◆ 具身智能数采与数据行业的商业化突围，依赖于数据服务商、整机/应用企业、政府端三侧的协同共进，在高价值垂直场景中，将高质量数据转化为不可复制的场景理解与运营优势，最终实现产业的整体突破与价值兑现。
- ◆ 数据服务商从数据供应商升级为**基础设施和合作伙伴**，实现数据集标准化复售/订阅、数据清洗与定制化采集、模型调优与评测服务等多元模式的盈利模式；整机/应用企业逐渐成为**数据闭环**的发起者与最终价值的实现者，除整机销售外，实现服务订阅、数据资产变现、生态分成的盈利模式；政府端成为生态构建师，为产业健康、安全、快速发展提供土壤，最终获得产业主导权、税收、就业和高科技产业集群的综合回报。

### 数据服务商

将原始数据转化为可驱动模型性能提升的标准化产品与服务

**深耕垂直场景**：锁定1-2个高潜、高价值赛道，与头部企业绑定，共建场景数据标准；

**构建数据混合生成能力**：合成数据+真实数据灵活配比，解决高质量数据稀缺和成本问题；

**产品化**：打造可复用的数据工具链与标注平台，把非标服务变成标准化订阅产品；

成为数据基础设施与生态伙伴

盈利：数据集标准化复售/订阅、数据清洗与定制化服务、模型调优与评测服务。

### 整机/应用企业

利用自身场景优势，构建不可复制的数据壁垒

**垂直闭环**：选最短商业化路径的场景，跑通“部署→运行→数据回流→迭代”的最小数据飞轮；

**数据内置**：把采集、回传、处理做成产品核心模块，持续低成本获取真实交互数据；

**生态开放**：有限开放数据接口，吸引伙伴共建应用，丰富场景能力；

成为平台型场景运营商

盈利：整机销售、服务订阅收入、数据资产变现、生态分成收入等多元化收入来源。

### 政府端

为产业健康、安全、快速发展创造土壤

**场景开放**：打造具身智能示范应用试点，为企业提供试验场和高质量资料来源；

**软性基础设施建设**：建设国家级数据集、评测基准和仿真平台，降低全行业研发门槛；

**标准共建**：推动产学研联合制定行业标准，打破数据孤岛，规避未来治理风险；

成为“生态构建师”

盈利：就业、税收、产业集群、产业主导权。

## 4.2 未来趋势

- ◆ 具身智能数采与数据产业正在迈向“工业化大生产”的数据基础设施阶段。未来，数据成为差异化竞争的核心变量，真机、无本体、仿真数据走向深度协同，共同推动行业走向标准化、平台化、生态化。未来竞争重点也会发生变化，从数据量转向有效信息密度、有效内容发展。垂类场景数据将成为护城河。



### 竞争范式：数据军备竞赛

#### 数据成为差异化竞争的核心变量

数据积累：谁先完成数百万小时来自真实实业的物理互动数据积累，谁就率先进入下一级



### 数据供给：互补共生

#### 真机、无本体、仿真/合成数据成为分层协同的互补关系

真实数据负责落地：提供物理真实性锚点

合成数据负责扩量：覆盖长尾场景

无本体数据负责连接：成为规模与精度之间可拓展的中间态



### 数据基础：标准、平台、生态三位一体

#### 标准化加速、平台化崛起、生态化闭环

标准化：硬件格式、数据格式、评估基准、安全伦理等更加规范，有标准可依

平台化：公共训练场/数据工厂互联互通，云边端架构协同、数据即服务（向订阅制演进）

生态化：数据-产品-应用形成闭环



### 数据价值：有效信息密度成为核心指标

#### 从量到效，数据价值度量为有效信息密度

有效性：单位数据带来的模型性能提升

质量：保证分布一致性，自动化质检

成本降低：由人力密集型转向技术密集型，边际效应凸显



### 垂直场景：数据走向专用

#### 处置场景深化，数据从通用走向专用，公共数据专业化

通用数据：支撑基础能力

垂类数据：建立场景壁垒，如工业、物流、商业等

专用数据：实现极致性能，精密操作、高危作业、医疗康养等

## 4.3 战略建议

- ◆ 具身智能数采与数据行业的竞争本质正在转向**数据获取效率与转化效率**的双重比拼。谁能高效、低成本地获取并利用高质量数据，谁就掌握了产业竞争的主动权与定价权。未来，将以“**混合动力**”降本增效，通过垂直场景的**数据飞轮**实现价值闭环与商业落地，利用生态标准卡位是通往成功的必由之路。当数据采集从“高成本、定制化、人力密集型”转向“**低成本、标准化、自动化**”时，产业才真正进入“**数据驱动**”阶段。

### 亿欧智库：具身智能数采与数据行业战略建议

#### 构建“混合动力”数据供应链

- **仿真优先**：建高保真仿真环境，批量生成成长尾场景合成数据，用于预训练和安全测试，降低真机试错成本。
- **精准采真**：聚焦关键稀缺场景，通过遥操作、穿戴设备等方式精准采集真机交互数据，保障落地可靠性。
- **动态配比**：按任务需求和模型阶段，灵活调配仿真、视频、动捕与真机数据比例，实现最优性价比。

#### 深耕垂直场景，跑通“数据飞轮”

- **精准切入**：选1-2个需求明确、边界清晰、ROI可算的高价值垂直场景。
- **部署即采集**：设备进场同时建数据回传通道，边解决问题边积累场景数据资产。
- **增长闭环**：用回流数据持续迭代模型，以效果提升推动同场景规模化复制。



#### 推动生态协同，抢占标准先机

- **主导标准**：参与数据格式、标注规范、接口协议及评测基准等行业标准制定，抢占话语权。
- **开放生态**：建立数据共享交换机制，汇聚多元数据集，降低全行业获取成本。
- **抢抓政策**：积极申报国家级/行业级数据标注基地、高质量数据集等专项，借力公共资源扩大规模与权威性。

- ◆ 亿欧智库经过桌面研究，结合相关公开报道及对相关企业、专家访谈后作出此份报告。在此，亿欧智库感谢相关企业及业内专家的鼎力支持。
- ◆ 未来，亿欧智库将持续密切关注人工智能领域，通过对于行业的深度观察，持续输出更多有价值的研究成果，助力产业可持续创新发展。欢迎报道读者与我们交流联系，提出报告建议。
- ◆ 特别鸣谢

amigos



## ◆ 团队介绍:

亿欧智库 (EOIntelligence) 是亿欧旗下的研究与咨询机构。为全球企业和政府决策者提供行业研究、投资分析和创新咨询服务。亿欧智库对前沿领域保持着敏锐的洞察，具有独创的方法论和模型，服务能力和质量获得客户的广泛认可。

亿欧智库长期深耕新科技、消费、大健康、汽车出行、产业/工业、金融、碳中和等领域，旗下近100名分析师均毕业于名校，绝大多数具有丰富的从业经验；亿欧智库是中国极少数能同时生产中英文深度分析和专业报告的机构，分析师的研究成果和洞察经常被全球顶级媒体采访和引用。

以专业为本，借助亿欧网和亿欧国际网站的传播优势，亿欧智库的研究成果在影响力上往往数倍于同行。同时，亿欧内部拥有一个由数万名科技和产业高端专家构成的资源库，使亿欧智库的研究和咨询有强大支撑，更具洞察性和落地性。

## ◆ 报告作者:



王思雨

亿欧智库分析师

Email: wangsiyu@iyiou.com



周慧慧

亿欧智库 分析师

Email: zhouhuihui@iyiou.com

## ◆ 报告审核:



孙毅颂

亿欧智库研究总监

Email: sunyisong@iyiou.com

## ◆ 版权声明:

本报告所采用的数据均来自合规渠道，分析逻辑基于智库的专业理解，清晰准确地反映了作者的研究观点。本报告仅在相关法律许可的情况下发放，并仅为提供信息而发放，概不构成任何广告。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议。本报告的信息来源于已公开的资料，亿欧智库对该等信息的准确性、完整性或可靠性作尽可能的追求但不作任何保证。本报告所载的资料、意见及推测仅反映亿欧智库于发布本报告当日之前的判断，在不同时期，亿欧智库可发出与本报告所载资料、意见及推测不一致的报告。亿欧智库不保证本报告所含信息保持在最新状态。同时，亿欧智库对本报告所含信息可在不发出通知的情形下做出修改，读者可自行关注相应的更新或修改。

本报告版权属于亿欧智库，欢迎因研究需要引用本报告内容，引用时需注明出处为“亿欧智库”。对于未注明来源的引用、盗用、篡改以及其他侵犯亿欧智库著作权的商业行为，亿欧智库将保留追究其法律责任的权利。

## ◆ 关于我们:

亿欧是一家专注科技+产业+投资的信息平台和智库；成立于2014年2月，总部位于北京，在上海、深圳、南京、纽约设有分公司。亿欧立足中国、影响全球，用户/客户覆盖超过50个国家或地区。

亿欧旗下的产品和服务包括：信息平台亿欧网（[iyiou.com](http://iyiou.com)）、亿欧国际站（[EqualOcean.com](http://EqualOcean.com)）、研究和咨询服务亿欧智库（EOIntelligence），产业和投融资数据产品亿欧数据（EOData）；行业垂直子公司亿欧大健康（EOHealthcare）和亿欧汽车（EOAuto）等。

◆ 基于自身的研究和咨询能力，同时借助亿欧网和亿欧国际网站的传播优势；亿欧为创业公司、大型企业、政府机构、机构投资者等客户类型提供有针对性的服务。

## ◆ 创业公司

亿欧旗下的亿欧网和亿欧国际站是创业创新领域的知名信息平台，是各类VC机构、产业基金、创业者和政府产业部门重点关注的平台。创业公司被亿欧网和亿欧国际站报道后，能获得巨大的品牌曝光，有利于降低融资过程中的解释成本；同时，对于吸引上下游合作伙伴及招募人才有积极作用。对于优质的创业公司，还可以作为案例纳入亿欧智库的相关报告，树立权威的行业地位。

## ◆ 大型企业

凭借对科技+产业+投资的深刻理解，亿欧除了为一些大型企业提供品牌服务外，更多地基于自身的研究能力和第三方视角，为大型企业提供行业研究、用户研究、投资分析和创新咨询等服务。同时，亿欧有实时更新的产业数据库和广泛的链接能力，能为大型企业进行产品落地和布局生态提供支持。

## ◆ 政府机构

针对政府类客户，亿欧提供四类服务：一是针对政府重点关注的领域提供产业情报，梳理特定产业在国内外的动态和前沿趋势，为相关政府领导提供智库外脑。二是根据政府的要求，组织相关产业的代表性企业和政府机构沟通交流，探讨合作机会；三是针对政府机构和旗下的产业园区，提供有针对性的产业培训，提升行业认知、提高招商和服务域内企业的水平；四是辅助政府机构做产业规划。

## ◆ 机构投资者

亿欧除了有强大的分析师团队外，另外有一个超过15000名专家的资源库；能为机构投资者提供专家咨询、和标的调研服务，减少投资过程中的信息不对称，做出正确的投资决策。

© 亿欧智库

© 亿欧智库

© 亿欧智库



扫码关注亿欧智库  
查看更多研究报告



扫码添加小助手  
加入行业交流群

 亿欧智库

网址: <https://www.iyiou.com/research>

邮箱: [hezuo@iyiou.com](mailto:hezuo@iyiou.com)

电话: 010-53321289

© 亿欧智库

© 亿欧智库

© 亿欧智库