



2025 AI+ Development
Digital Summit

AI+ 研发数字峰会

拥抱AI 重塑研发

05/23-24 | 上海站



2025 AI+研发数字峰会

拥抱AI 重塑研发 AI+ Development Digital Summit

下一站预告

08/08-09 | 北京站

11/14-15 | 深圳站



查看会议详情

北京站论坛设置

大模型和 AI 应用评测

智能存储与检索技术

下一代知识工程

AI+ 金融业务创新

智能需求工程

智能体与研发效率工具

AI 产品运营与出海策略

大模型安全与对齐

大模型应用开发框架与实践

智能体经济 (Agentic Economy)

智能测试工具的开发与应用

具身智能与机器人

代码生成及其改进

AI+ 新能源汽车

AI 前沿技术探索与实践

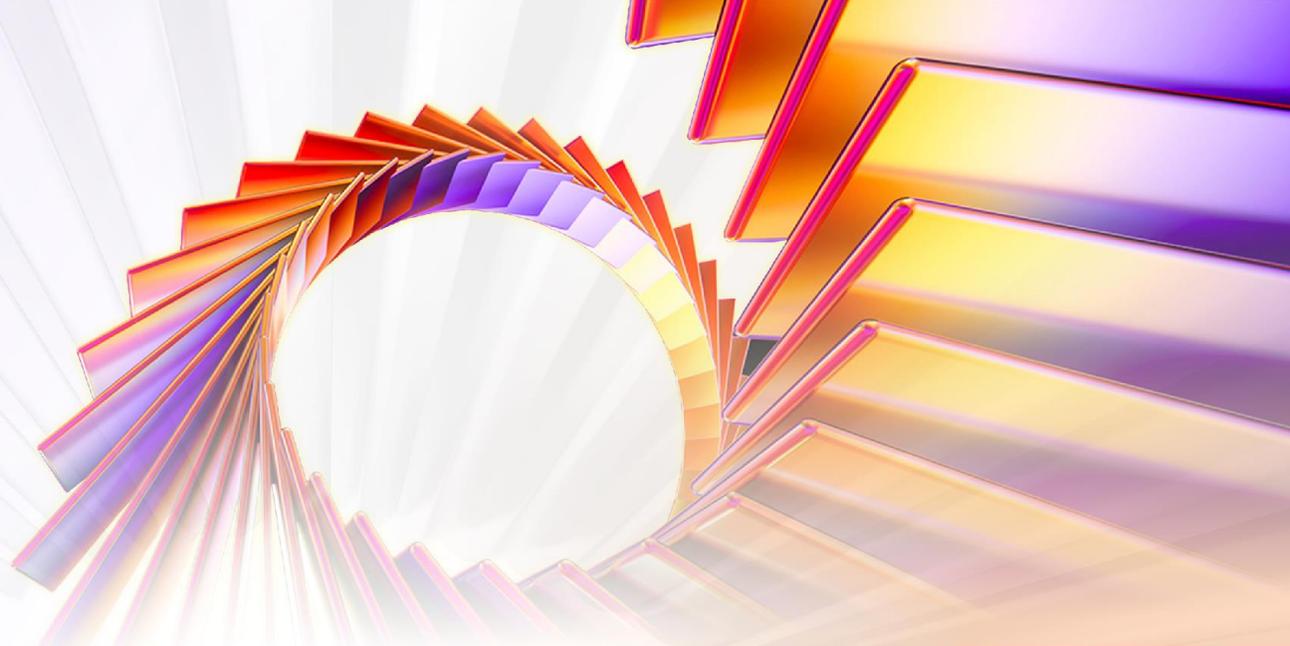


| 05/23-24 | 上海站

2025 AI+ Development
Digital Summit

AI+研发数字峰会

拥抱AI 重塑研发



基于MobileFlow的 终端自动化测试效能跃升实践

唐晓璇 | 蚂蚁集团



唐晓璇

蚂蚁集团-CIO MobileFlow 算法负责人

就职于蚂蚁集团，专注于研发效能优化方向，主要工作方向为AI驱动的研发效能体系创新。作为MobileFlow技术团队的算法工程师，聚焦多模态大模型与智能体技术的前沿探索，主要研究方向包括：1) 视觉-语言联合建模在UI自动化测试中的精准指令理解与执行；2) LLM-based Agent的复杂任务分解与动态规划能力优化；3) 跨平台界面操作智能体的端到端自动化测试框架构建。为复杂场景下的研发效率与质量提升提供技术驱动力。

目录

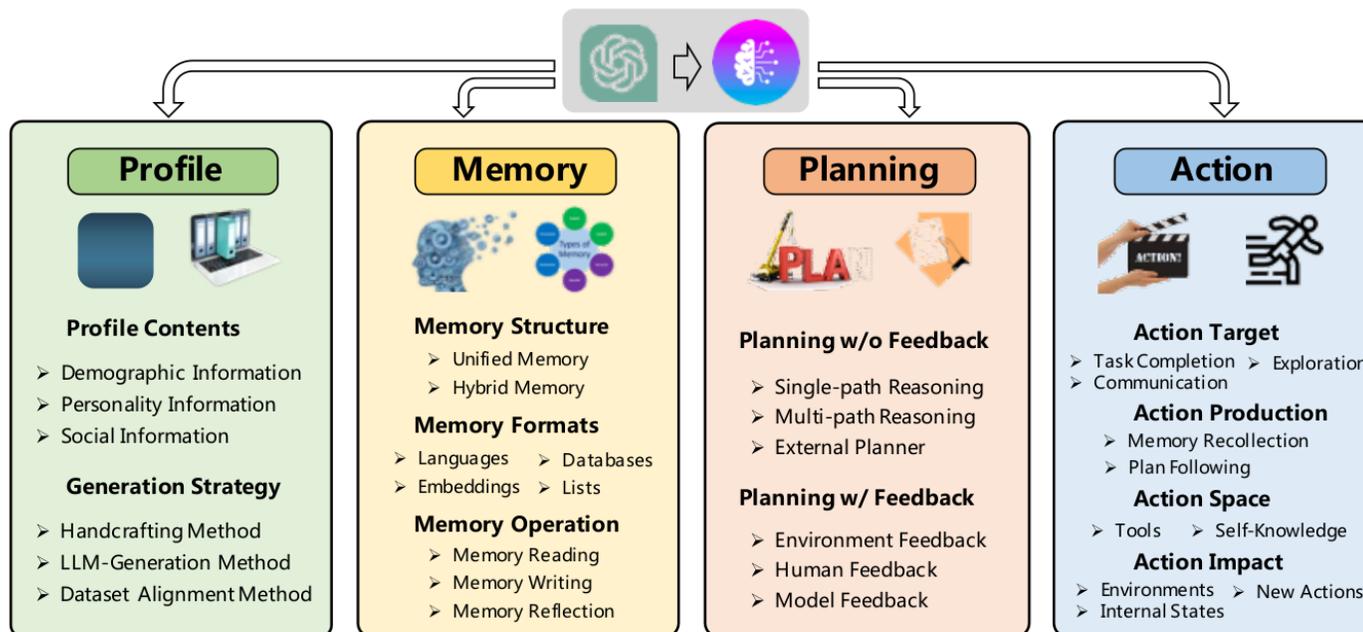
CONTENTS

- I. 智能体发展的背景与趋势
- II. 智能体评测通用痛点
- III. 蚂蚁MobileFlow平台介绍
- IV. MobileFlow平台多模态技术介绍
- V. 工业级场景落地与应用
- VI. 思考与展望

PART 01

智能体发展的背景与趋势

什么是基于大语言模型的智能体 (LLM-based Agent) ?



- ✓ **LLM 核心**
任务规划, 推理, 文本生成
- ✓ **记忆模块**
存储历史状态和经验
- ✓ **工具集:**
通过预定义接口调用外部工具 (如数据库、API、代码执行器)
- ✓ **执行模块:**
执行具体动作

基于大语言模型的智能体是一种能够自主感知环境、规划决策并执行动作的系统，其核心能力由大语言模型驱动。这类智能体通过整合语言理解、推理、记忆和交互能力，模拟人类在复杂场景中的行为逻辑，实现从任务目标到实际操作的全链路闭环。



怎么查询医保还有多少钱？



自然的多模态交互

- 无缝接收用户的相关输入（如语音、文本、图像、生理信号等），构建完整的上下文感知能力

需求感知理解

- 意图理解，需求分类，上下文感知，

自主任务编排拆解

- 将复杂目标分解为可执行的子任务并动态调整计划

工具执行并返回结果

- 自动化工具执行，并以用户友好的方式反馈结果（如可视化图表、语音总结）



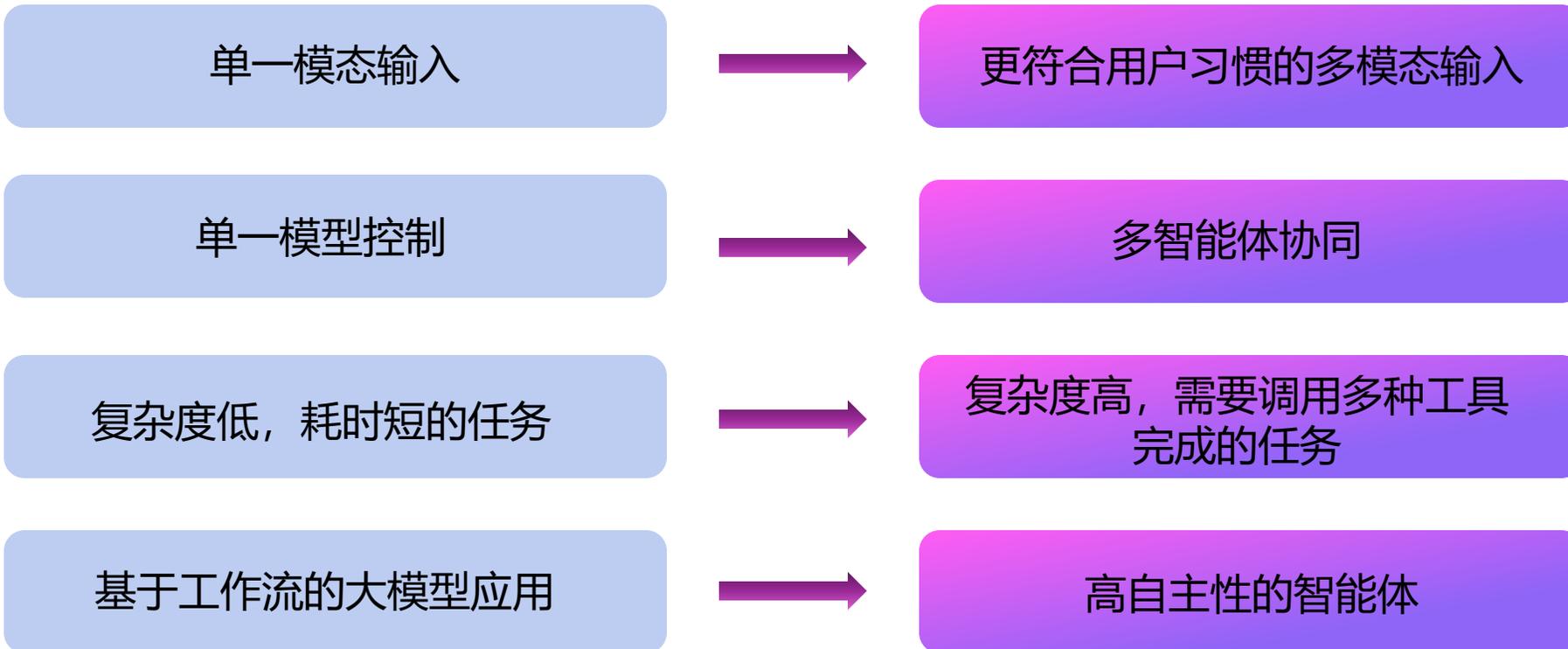
智能体常用场景



在大模型和智能体迅速发展的背景下，智能体评测的方式也在不断进化。传统的评测方法已经难以满足移动端智能体和多轮对话系统的复杂需求，智能体评测升级成为必然的需求。



智能体发展趋势



PART 02

智能体应用评测通用痛点

痛点一

缺乏真实场景模拟

- 多在理想或简化的条件下测试，无法完全模拟真实的移动端使用和用户行为。例如节假日第三方发起的营销活动等场景很难通过人工有效覆盖，或者直接导致评测脚本执行报错。



千人千面的内容推荐

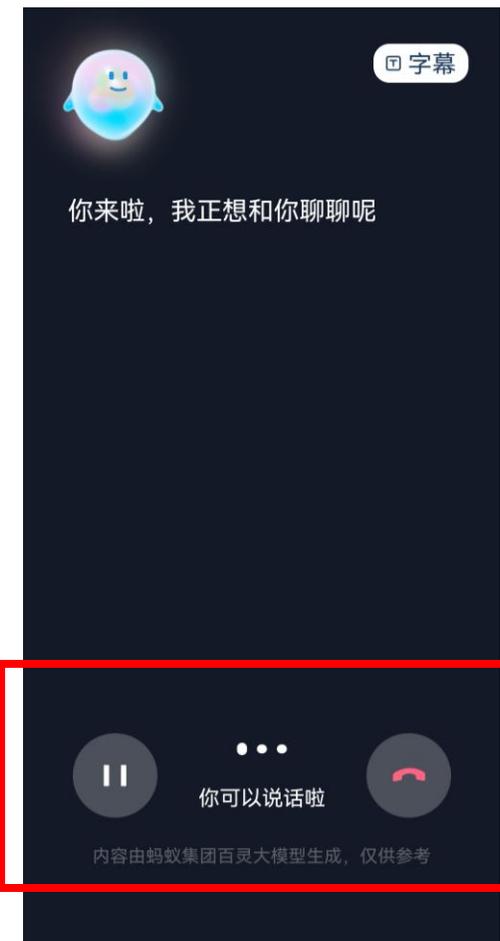


不同用户不同行为特征

痛点二

多模态交互支持差

- 传统方式使用精准文字或图片匹配作为脚本规则路由方式。无法有效评测图片、语音等多模态信息的理解和处理。



多种模态的输入输出

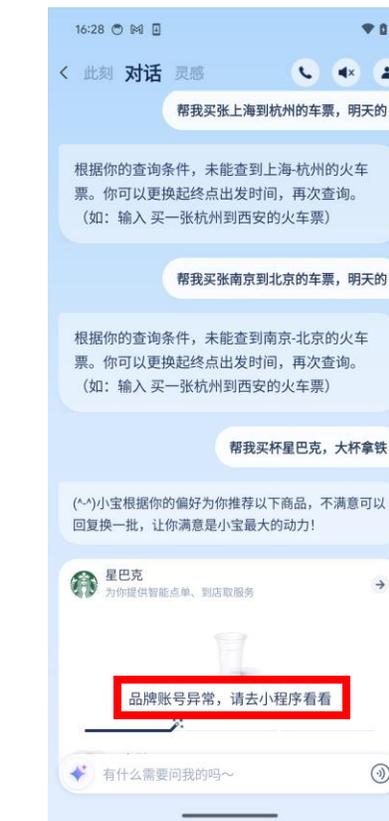
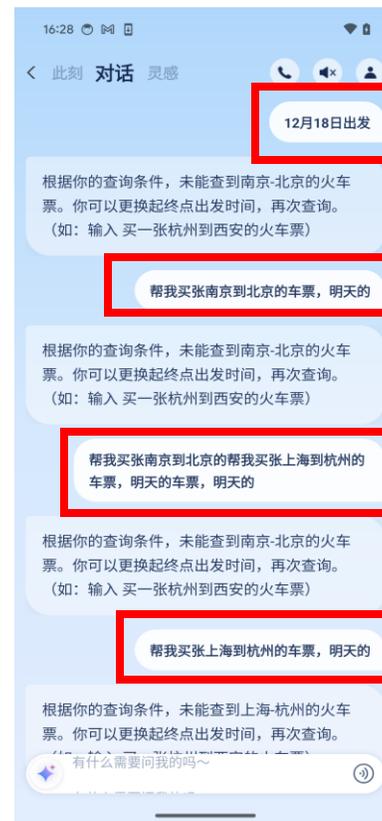
痛点三

评测范围有限

- 因为待评测模型产出存在不确定性，当扩展到多轮对话交互后，使用传统RPA + 规则脚本驱动的方式覆盖成本成指数级增加，并且执行稳定性差、成功率低。实际还需要依靠人工兜底，评测效率低。



无法穷尽的多轮对话

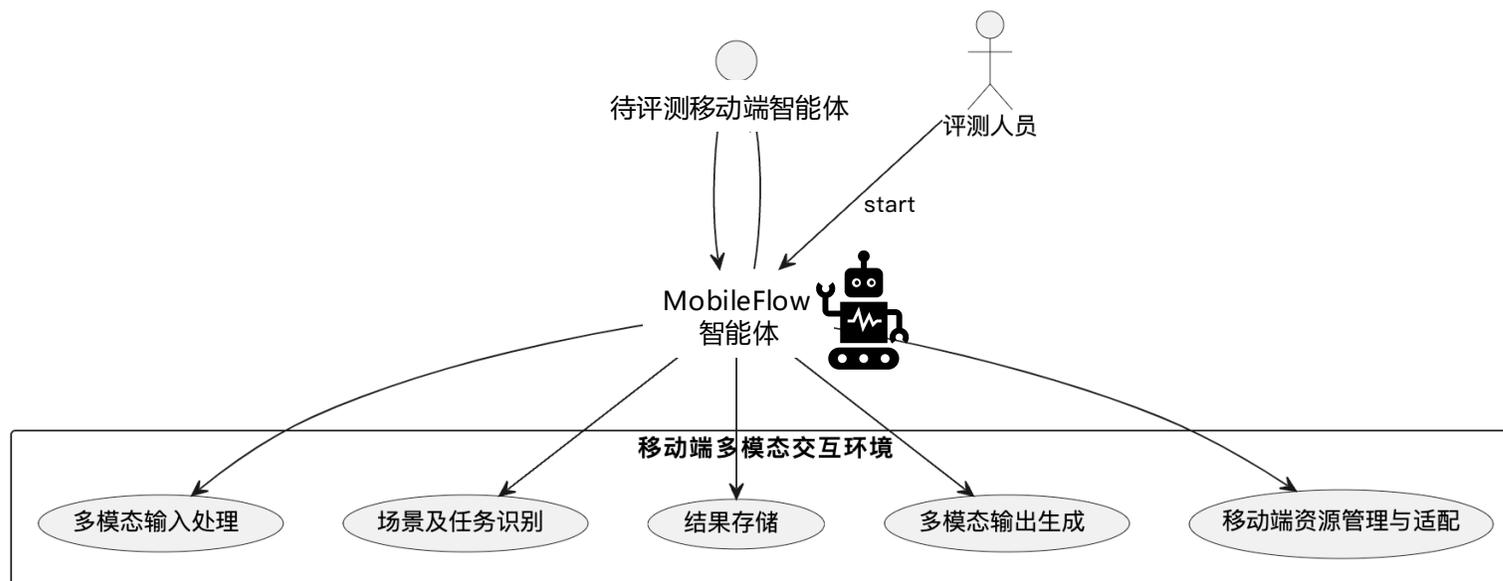


五花八门的报错信息

PART 03

蚂蚁MobileFlow平台介绍

MobileFlow 平台能力



智能体评测痛点

缺乏真实场景模拟

多模态支持差

评测范围有限

MobileFlow平台能力

动态交互+真实用户模拟

跨模态协同能力

测试动线自动化生成

MobileFlow 平台架构



支持业务场景

20+ App/业务接入

5K+ 在线终端设备

130M+ 年自动化任务

2.5K+ 月活用户

300K+ 自动化脚本

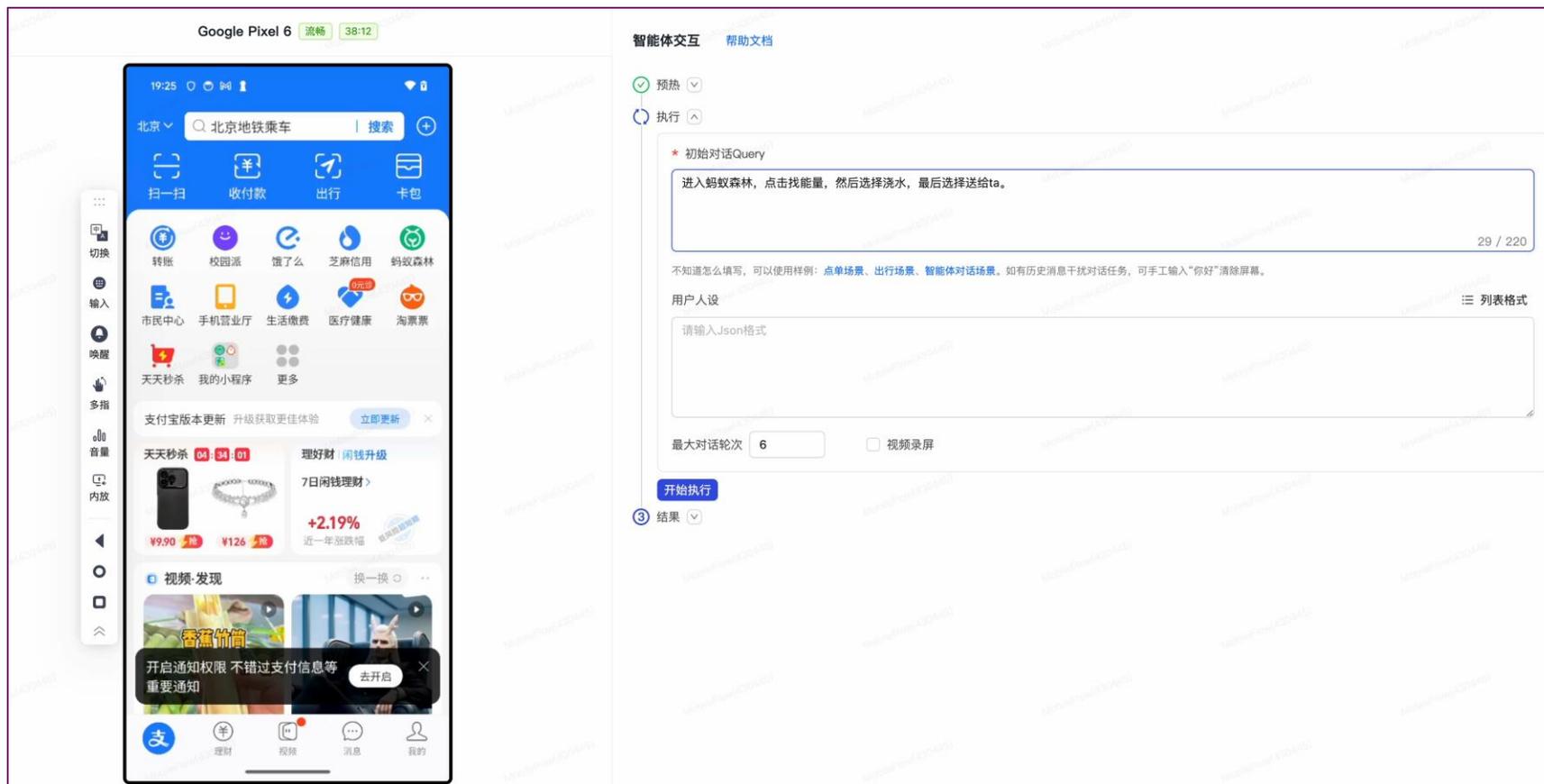
多模态算法技术

系统平台能力

云真机资源



MobileFlow平台介绍 – 交互演示Demo



平台demo演示:
进入蚂蚁森林, 点击找能量, 然后选择浇水, 最后选择送给ta。

- ✓ MobileFlow智能体动态环境交互
- ✓ 指定人设, 模拟真实用户行为



- ✓ 大小模型协同
- ✓ 多模态大模型+语言大模型多种模态交互



云真机：远程操控云端手机

功能：

线上问题复现

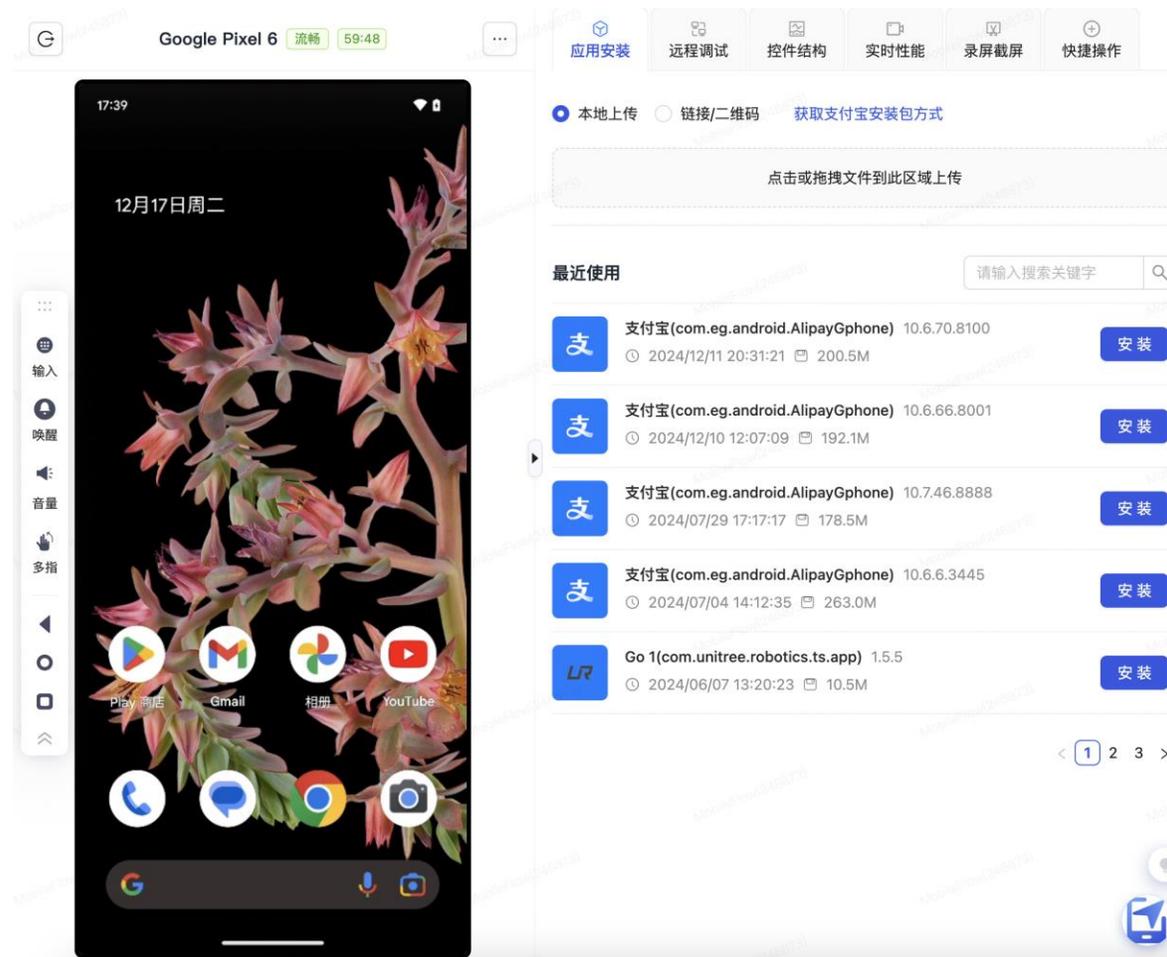
日志抓取，Mocking，收集性能问题

UI效果预览

支持多种型号，多种操作系统设备，产品覆盖全

便捷端云协同联调

支持远程仿真环境联调，把手机拉取到本地调试



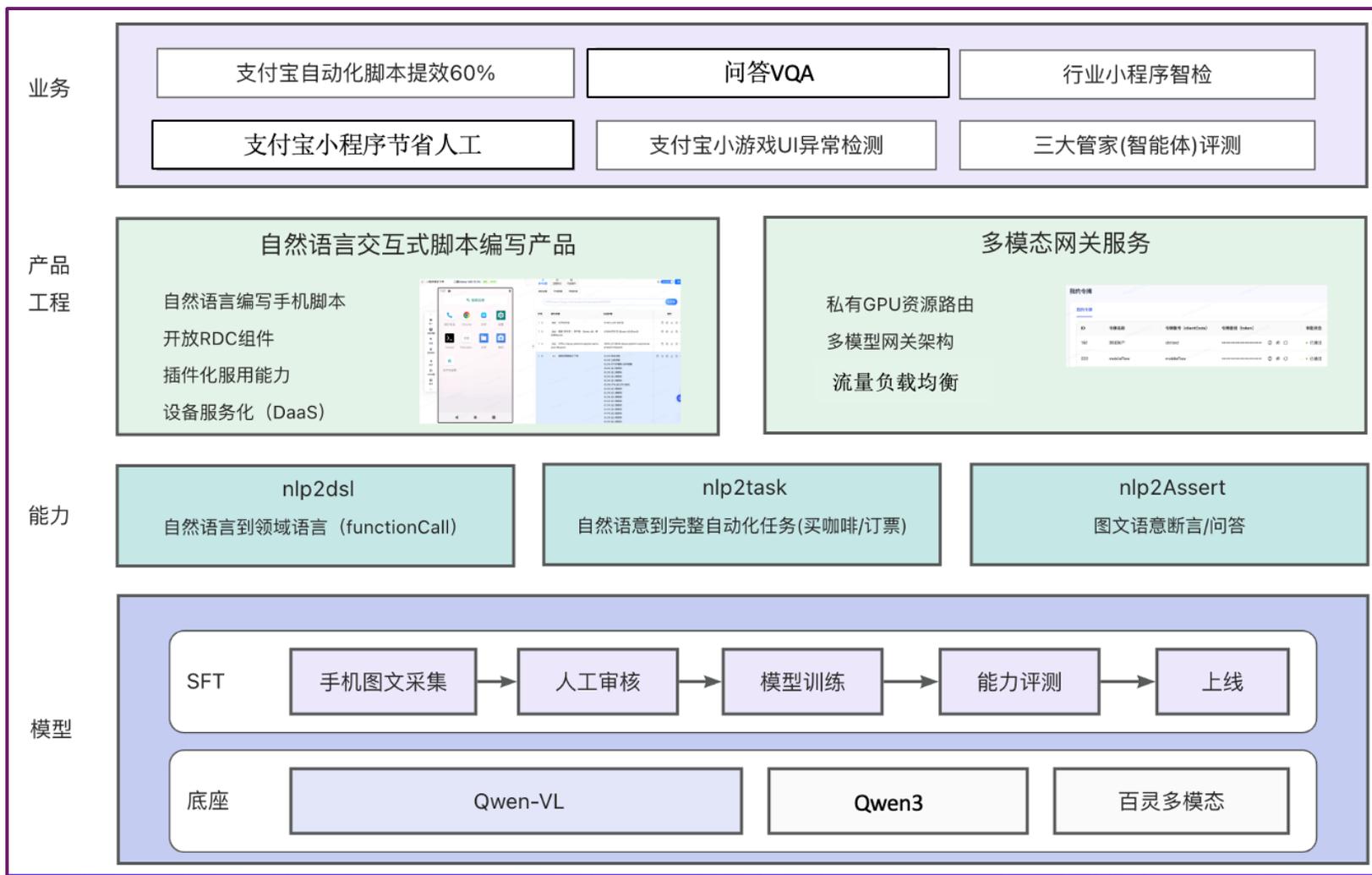
- ✓ 真实场景模拟
- ✓ 测试任务自动化执行



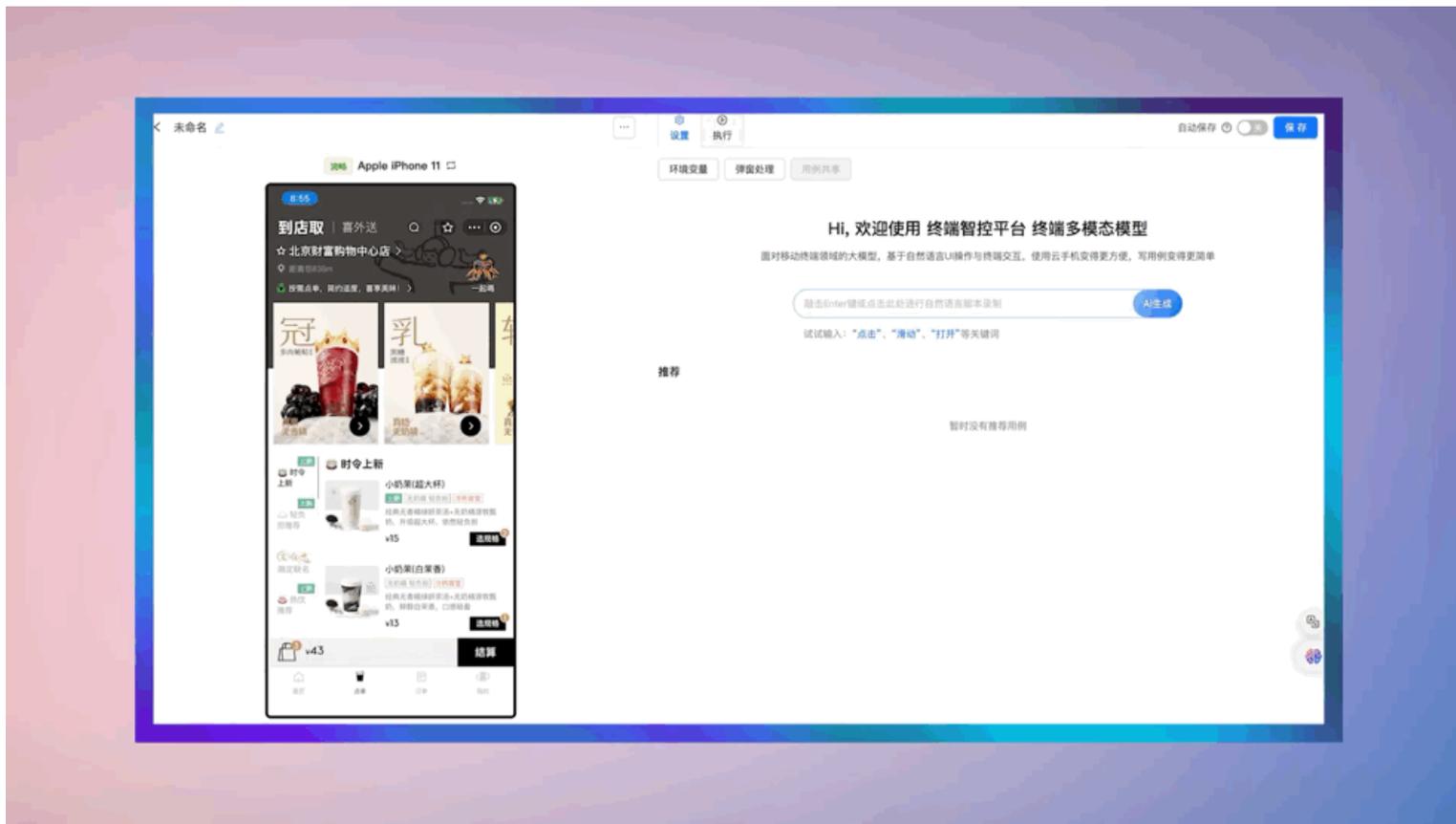
PART 04

MobileFlow平台多模态技术介绍

MobileFlow – 多模态大模型驱动测试脚本生成



▶ MobileFlow – 多模态大模型驱动测试脚本生成



用户输入：
选择任意商品并进行下
单。

自然语言生成脚本



MobileFlow - 多模态大模型驱动测试脚本生成

在外卖中，帮我下单两杯茉莉手打柠檬茶

从外卖点单进入点单页面

{“name”:
“click”, “point”:
[397, 706]}



选择下单地址

{“name”:
“click”,
“point”:
[87, 406]}



选择商品

{“name”:
“click”,
“point”:
[780,
996]}



选择数量

{“name”:
“click”,
“point”:
[768, 2001]}

加入购物车

{“name”:
“click”,
“point”:
[780,
1997]}



点击去结算

{“name”:
“click”,
“point”:
[732,
1890]}



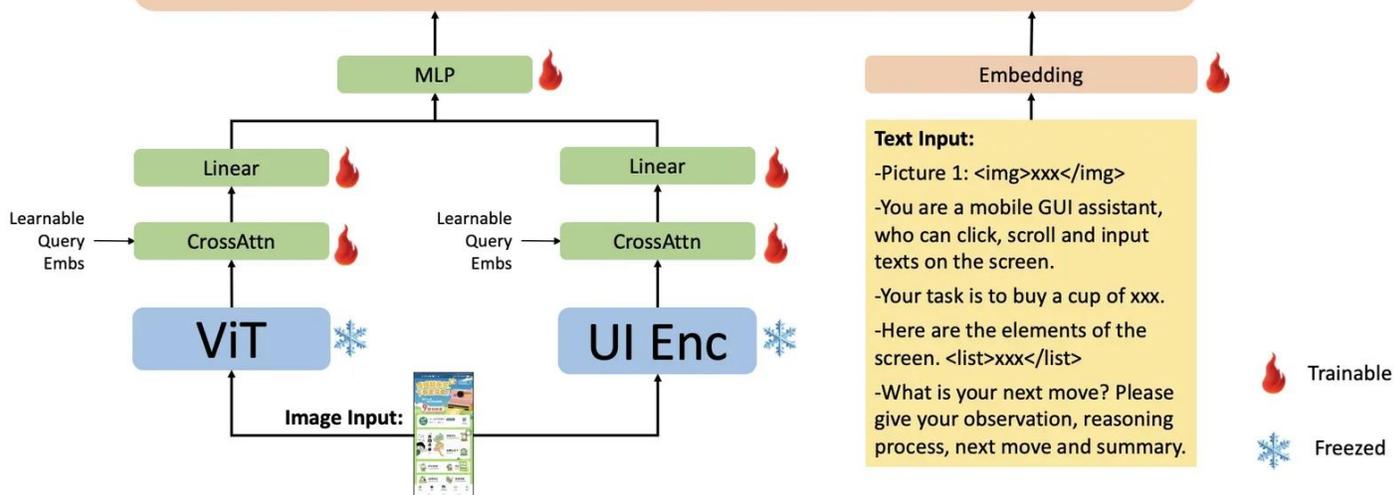
Finish

MobileFlow- 模型结构

Text Output: Observation: In the picture ... Reasoning: Considering the elements ... Move: Click ... Summary: ...

Large Language Model

(MoE expanded)



模型结构

完全基于图像的Native GUI Agent Model
含三个主要组成部分：

- ✓ 视觉编码器、
 - ✓ 视觉-语言适配器
 - ✓ 大型语言模型
-
- ✓ **混合视觉编码**：理解页面布局结构、控件的手机UI场景视觉理解编码器
 - ✓ **视觉语言对齐**：网络结构优化 + 多任务训练，强化大模型跨模态任务能力

Training:

- GUI Grounding
- Visual Question Answering
- UI Element Captioning
- Real Component Generation



▶ MobileFlow – 模型效果评测

Method	Metrics	Different business areas of testing data						All
		Food Delivery	Food Walkin	Medical Service	Fund Select	Insurance	Gaming	
GPT-4v	WTSR	0.1833	0.1876	0.1138	0.1428	0.2021	0.4132	0.2071
	SSR	0.5716	0.4647	0.3571	0.4857	0.5012	0.6613	0.5069
Qwen-vl-Max	WTSR	0.3650	0.2562	0.1643	0.2875	0.3076	0.5832	0.3273
	SSR	0.7338	0.7075	0.6962	0.5112	0.2425	0.7763	0.6113
Mobile-Flow	WTSR	0.4667	0.2917	0.3333	0.3810	0.3333	0.6363	0.4071
	SSR	0.8735	0.7921	0.8341	0.5353	0.6136	0.7199	0.7280

✓ 在真实手机UI场景测评，小样本SFT微调后场景能力超越GPT-4v等业界领先模型/方案





1. 弹窗识别与异常处理

利用专门训练的小模型来识别页面上的弹窗，并解决加载过程中出现的黑白屏等异常问题，确保动线执行过程中的稳定性和准确性。小模型辅助大模型应对终端环境的不稳定性（页面异常）

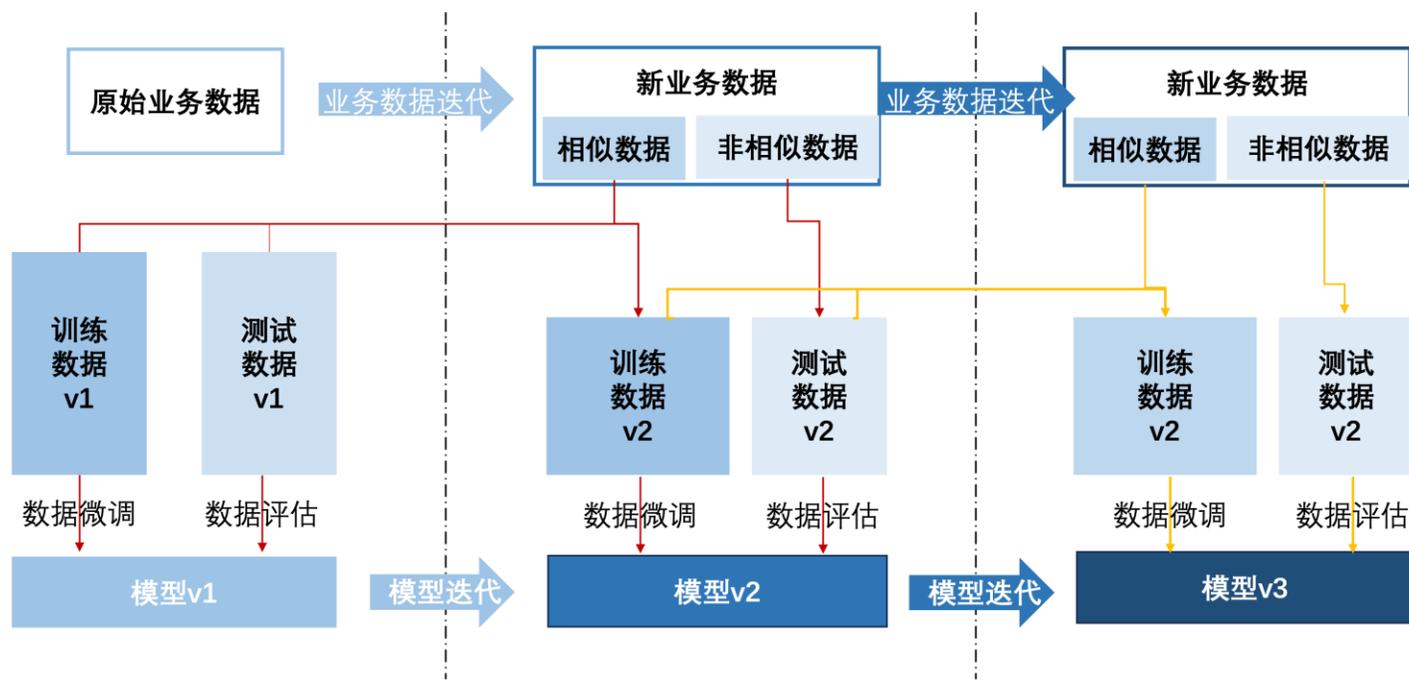


2. 控件分析辅助

通过控件识别的小模型提供额外的页面解析信息，确保大模型拥有充分的先验知识支持，补充开源多模态大模型在视觉上的不足，如对弹窗下视图、卡片式布局理解薄弱的问题。



MobileFlow模型训练数据飞轮



- ✓ 回归问题本质，通过大模型和人类意图能够良好对齐的特性，实现多样UI场景的兼容
- ✓ 以数据为核心，利用平台多年沉淀的UI自动化任务执行数据，构建手机场景多模态大模型训练数据
- ✓ 紧贴业务场景，大模型 + Agent + 产品的综合解决方案



欢迎使用 移动标注

在云真机上通过自然语言描述操作和屏幕交互方式，完成手机操作动线标注。标注数据将自动用于MobileFlow移动端大模型微调训练。

根据过往经验，新接入的业务场景（10步以内可完成动线）可以通过50条左右标注数据达到60%+动线准确率

业务合作 & 产品功能使用问题，请联系 [白堆](#)

开始标注

操作指引

数据标注平台



PART 05

工业级场景落地与应用

工业级场景落地与应用 – 支持支付宝质量管控提效



严格控制三端设备 (IOS, Android, 鸿蒙), 避免移植测试代码

一次操作, 同时控制三端设备完成相同动线操作, **降低测试门槛**



工业级场景落地与应用 - 鸣谢

- ✓ 2024年人工智能先锋案例
- ✓ MobileFlow 项目组专利证书4件，申请中14件
- ✓ 论文产出：6篇



- ✓ Mobileflow: A multimodal LLM for mobile GUI agent” The 38th Conference on Neural Information Processing Systems (NIPS 2024)
- ✓ MP-GUI: Modality Perception with MLLMs for GUI Understanding” (CVPR 2025)
- ✓ TestFlow: Advancing Mobile UI Testing through Multi-Step Reinforcement Learning (ISSTA Workshop 2025)
- ✓ PG-Agent: An Agent Powered by Page Graph (ACM MM 2025 submission)
- ✓ Towards an Inclusive MobileWeb: A Dataset and Framework for Focusability in UI Accessibility” (WWW 2025 submission)
- ✓ How Visually Impaired Users Penetrate the Accessibility of Mobile Apps?” (FES 2025 submission)



PART 06

思考与展望

设备集群化自动化

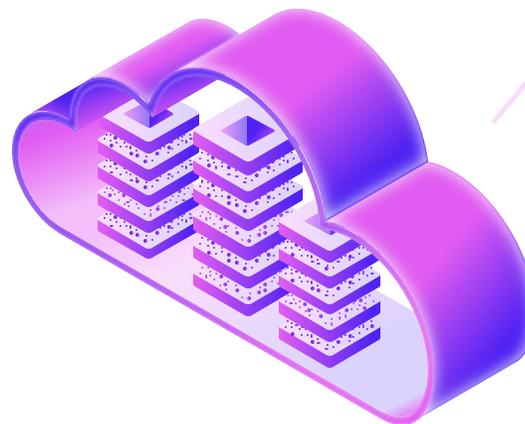
通过设备接入层统一接入IoT设备, 手机三端, PC等, 打造 DCP - Device Context Protocol

质量自动化

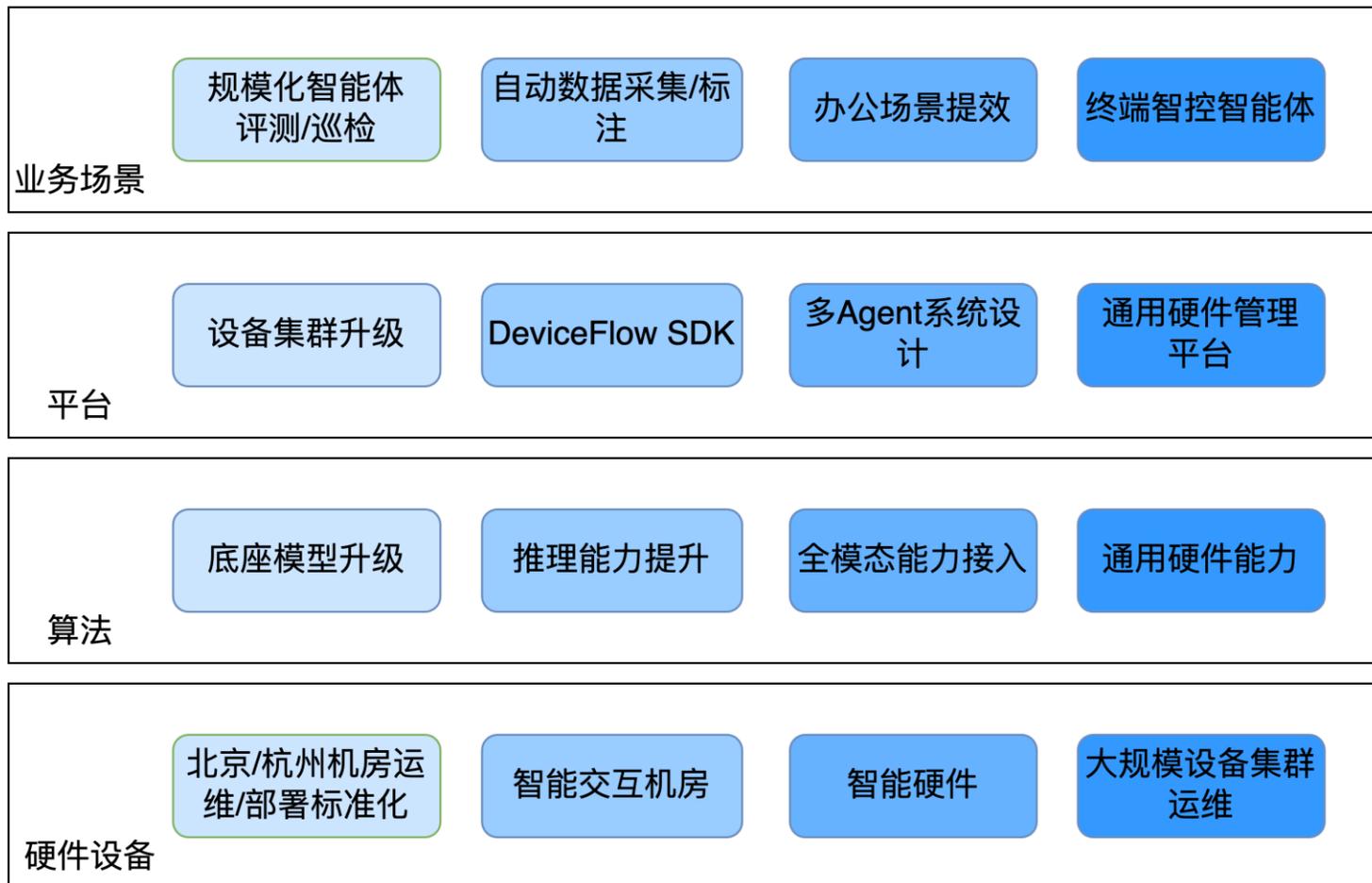
从RPA到APA, 用LLM自动化脚本生产

数据自动化

多设备协同, 数据采集自动化



思考与展望 - 打造终端智控智能体



- 未来将不仅是设备控制的“中枢神经”，更是智能化、生态化、人本化的超级入口。
- 通过技术深度整合、用户体验重构（无感交互+自进化）及生态开放协作，改变未来的人机交互范式。



参与调研您将优先获得



AiDD定制版
《AI+软件研发精选案例》



专属学习顾问
1对1需求对接

AiDD会后小调研

AiDD峰会致力于协助企业利用AI技术深化计算机对现实世界的理解，推动研发进入智能化和数字化的新时代。作为峰会的重要共建者，您的真知灼见对我们至关重要。衷心感谢您的参与支持！

2025 AI+研发数字峰会

拥抱 AI 重塑研发



扫码参与调研

科技生态圈峰会 + 深度研习

—1000+ 技术团队的选择



K+峰会 **敦煌站**
K+ 思考周®研习社
时间: 2025.08.29-30

K+峰会 **上海站**
K+ 金融专场
时间: 2025.09.26-27

K+峰会 **香港站**
K+ 思考周®研习社
时间: 2025.11.17-18



K+峰会详情



AIDD峰会 **上海站**
AI+研发数字峰会
时间: 2025.05.23-24

AIDD峰会 **北京站**
AI+研发数字峰会
时间: 2025.08.08-09

AIDD峰会 **深圳站**
AI+研发数字峰会
时间: 2025.11.14-15



AIDD峰会详情



2025 AI+研发数字峰会

AI+ Development Digital Summit

感谢聆听!

扫码领取会议PPT资料

