飞猪旅行

用AI操作GUI，飞猪以GUI Agent重塑流程自动化与交付效率

飞猪高级技术专家 梁筱武

飞猪旅行

极客邦科技 2026 年会议规划

促进软件开发及相关领域知识与创新的传播

参会咨询

查看会议

北京 · 1200人

QCon 全球软件开发大会

会议时间：4月16-18日

- Agentic Engineering
- AgentOps
- 下一代模型架构与推理优化
- AI 原生基础设施
- 知识工程实践
- AI 安全

深圳 · 1000人

AiCon 全球人工智能开发与应用大会

会议时间：8月21-22日

- Agentic AI
- 轻量化与高效推理
- 多模态应用
- AI + IoT 场景实践
- AI 工业化落地

北京 · 1000人

AiCon 全球人工智能开发与应用大会

会议时间：12月18-19日

- 大模型架构创新
- 多模态 AI 产业融合
- 具身智能
- AI for Science
- 大模型安全

4月 — 6月 — 8月 — 10月 — 12月

上海 · 1000人

AiCon 全球人工智能开发与应用大会

会议时间：6月26-27日

- AI Infra 系统工程
- 多 Agent 协作与实践
- 多模态融合
- 模型训练与推理创新
- 数据平台与特征服务

上海 · 1200人

QCon 全球软件开发大会

会议时间：10月22-24日

- AI Agent
- Vibe Coding
- 智能可观测
- 推理基建
- 模型攻防
- AI x 创造力

InfoQ 极客传媒

飞猪旅行

GUI自动化遇到的挑战与AI带来的新机会

01

GUI自动化遇到的挑战：界面变动即导致脚本失效，开发脚本效率低

飞猪旅行

控件依赖强&维护成本高

传统自动化依赖XPath或控件ID等定位元素，界面微调即导致脚本失效

每次界面变更需人工排查修复，平均耗时超2小时，迭代效率低下

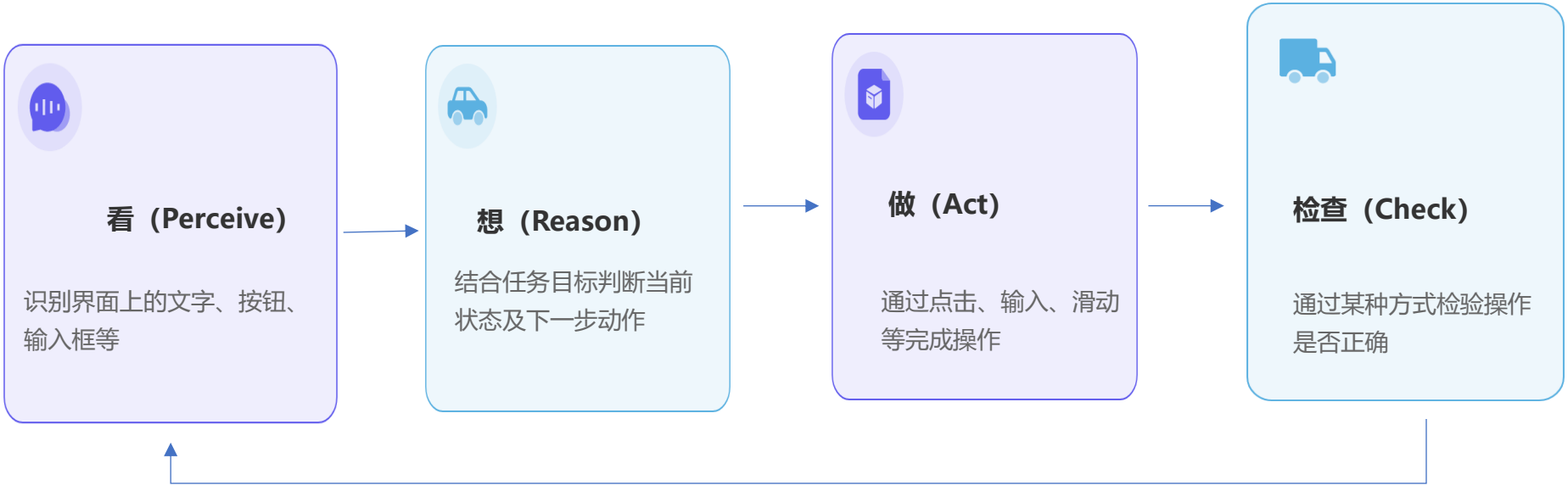
泛化能力弱

脚本无法跨端复用，同一业务在不同端需要重新开发脚本，资源浪费严重

智能性弱

纯静态，难以应对弹窗、网络异常、页面跳转等动态场景，缺乏环境感知能力

人类如何完成GUI操作：遵循 ‘感知-决策-执行-校验’ 闭环



GUI Agent：融合了多模态大模型、OCR辅助、决策自主规划与设备控制的端到端智能系统

GUI Agent vs GUI自动化：不依赖脚本，通过视觉+语义双重理解任务驱动GUI操作，具备强鲁棒性

飞猪旅行

摆脱控件依赖

基于屏幕视觉信息定位元素，无需控件ID或XPath，界面变更不影响执行稳定性

自然语言任务

结合OCR与多模态大模型，同时理解界面内容与用户自然语言意图，实现精准操作决策

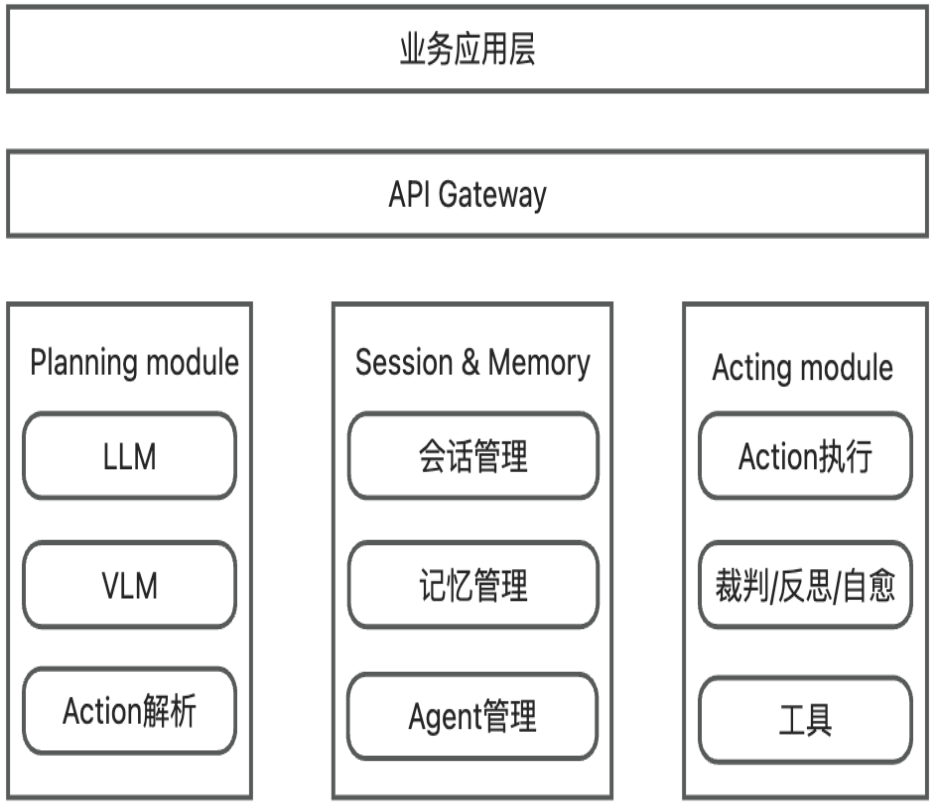
强鲁棒性能力

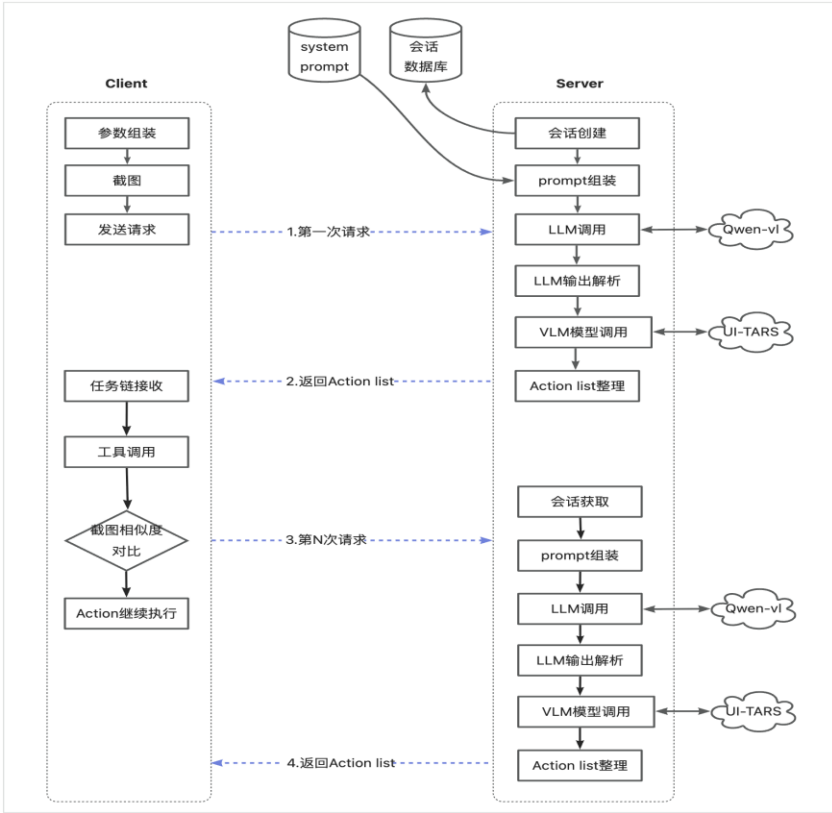
适应不同终端与动态场景，对弹窗、加载延迟等异常具备天然容错与恢复能力

GUI Agent技术架构

02

基于ReAct构建‘思考—行动—观察—迭代’的闭环流程





基于ReAct构建‘思考—行动—观察—迭代’的闭环流程

感知驱动决策

通过截图与OCR实时感知界面状态，结合用户指令（上下文）和历史记忆，由多模态大模型生成下一步动作意图

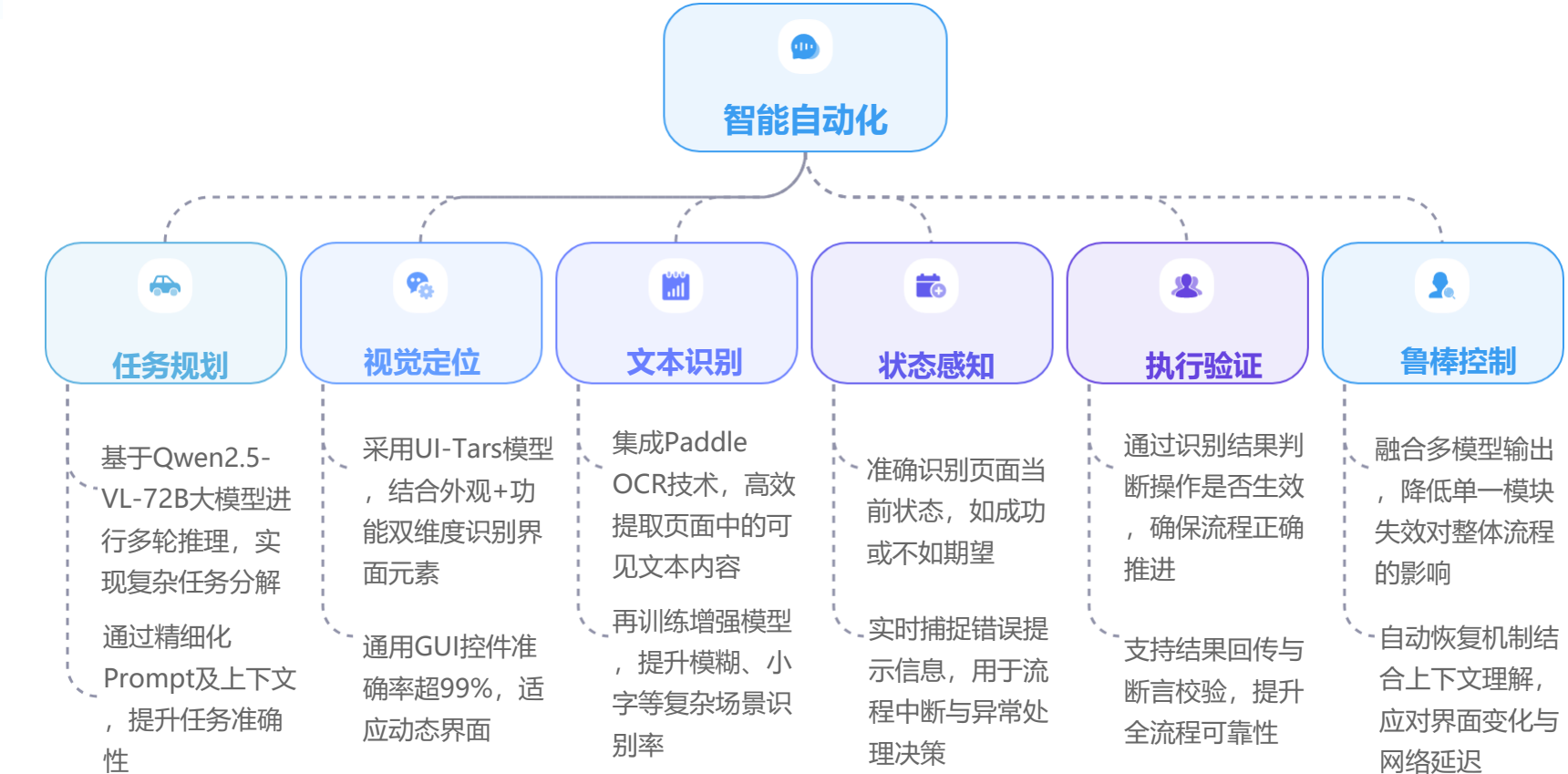
动作精准执行

规划引擎输出结构化操作指令，经UI-Tars模型定位坐标，通过ADB、WebDriver、WindowsDriver等在设备端完成点击、输入等操作

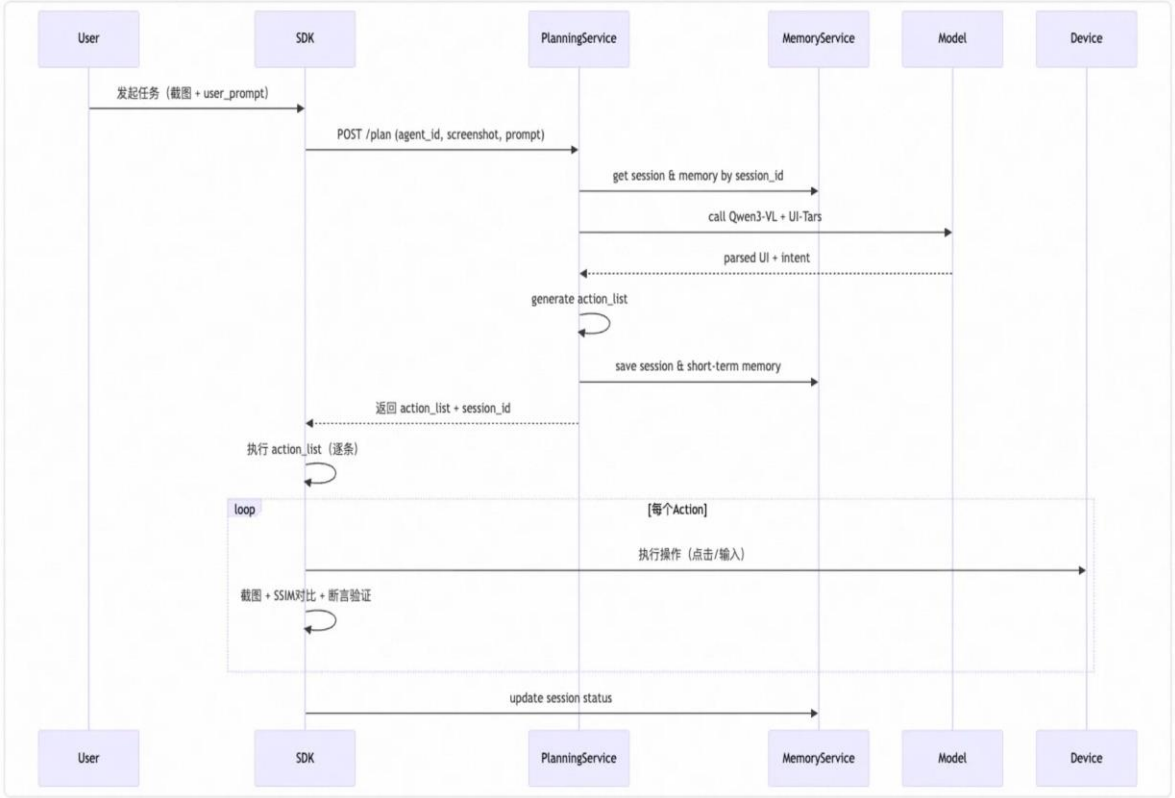
反馈闭环迭代

执行后自动截屏并验证结果，将新状态回传规划引擎，实现动态调整与试错式推进，直至任务完成

五大核心模块协同工作：任务规划、视觉定位、OCR辅助、记忆系统、客户端集成



五大核心模块协同工作：任务规划、视觉定位、OCR辅助、记忆系统、客户端集成



长期记忆

agent_session
-session_id
-user_id
-start_time
-status

agent_memory
-id
-session_id
-step
-thought
-action
-screenshot_url
-timestamp

GUI Agent关键技术

03

精细化Prompt设计、上下文增强与结构化动作空间，提升大模型理解准确性

飞猪旅行

角色明确化

在Prompt中明确定义Agent为‘GUI操作专家’，限定其行为边界与任务目标领域

上下文注入

1、融合历史动作与反馈结果
2、强调‘基于当前截图决策’避免过度预测未来状态
3、设备元信息（如platform: mobile）
4、终端可用动作集

少样本引导

嵌入典型任务的Few-shot示例，规范输出格式，引导模型生成符合执行要求的动作序列

动作结构化

定义标准化动作空间（name、description、parameters），要求元素描述包含功能、文案、位置三要素

动态坐标预测的成功率优化（grounding无法识别部分控件坐标）

飞猪旅行

相邻重复tought机制

对相邻重复thought更换模型或增加标记prompt

上下文注入追加超级指令prompt

通过超级指令强化此场景能力

四层容错与自愈机制有效处理弹窗、网络异常等动态干扰

飞猪旅行

L1：接口级重试

触发条件：HTTP 超时、限流等
机制：大模型调用失败自动重试 5 次

L2：流程级恢复

触发条件：关键词匹配
机制：OCR 检测到“系统繁忙”等后自动刷新重试

L3：语义级终止

触发条件：预设终止词
机制：出现 stopTexts（如“订单不存在”）则提前结束

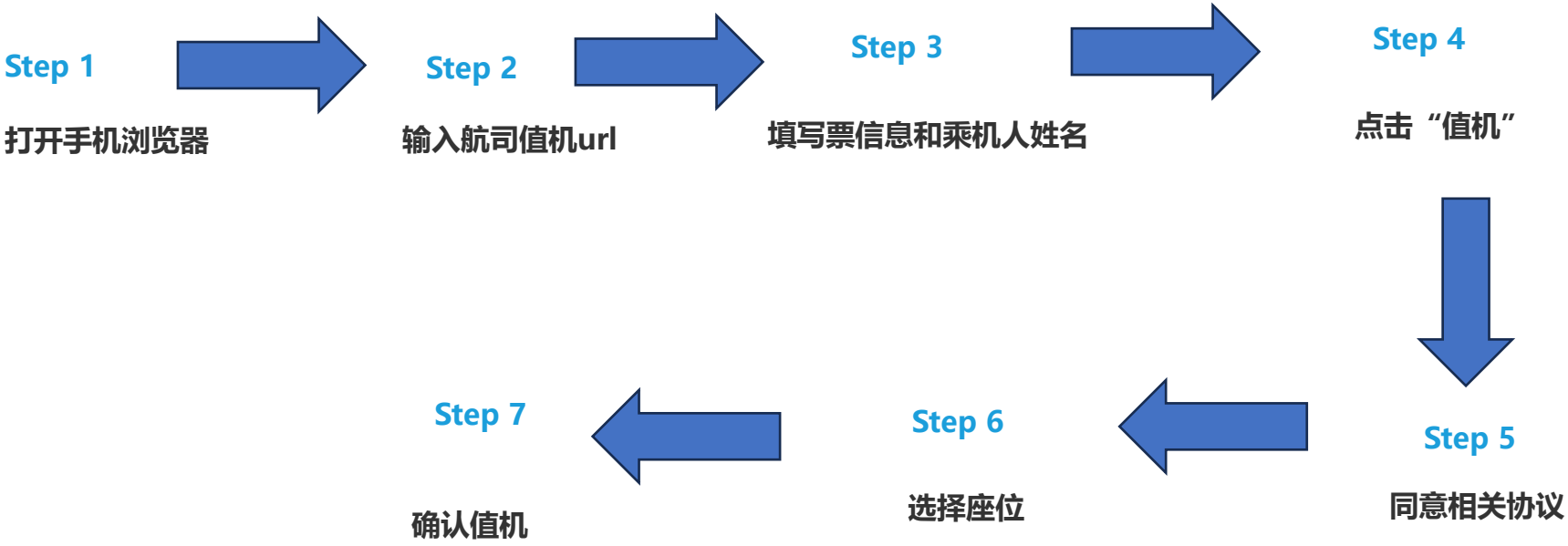
L4：可疑瘫痪级重启

触发条件：连续动作后但界面未变或变化巨大（可能跳转了页面）
机制：终止，重新启动规划

落地案例：航班值机流程自动化

04

传统GUI自动化：乘客‘张三’，票号‘FL123456’值机
7 Steps，Step by Step



GUI Agent：我是张三，票号是 FL123456，帮我完成值机

Agent自主完成‘进入订单页-搜索票号-填写信息-选座提交’全流程

任务启动

用户输入自然语言指令，Agent自动解析意图并规划初始动作路径

流程执行

Agent依次完成进入订单页、搜索票号、找到航班、输入乘机人信息填写与座位选择等操作

结果验证

通过OCR检测‘值机成功’字样，确认任务完成并返回结构化结果

端到端成功率95%以上， 端到端性能2分钟内，开发周期从3天缩短至0.5天

飞猪旅行

极简启动

仅需输入自然语言指令，无需编写代码，任务即可自动执行

全链路自治

Agent自主完成搜索、填写、选座到提交的完整操作流程

结果可验证

通过OCR识别‘值机成功’字样（包含上下文），确保操作结果真实可信

效率跃升

开发周期从3天降至0.5天，端到端成功率突破95%，端到端性能2分钟内

跨H5/小程序/App终端统一运行，真正实现一次定义、处处执行

飞猪旅行

统一技术栈

GUI Agent基于视觉与语义理解，屏蔽各终端差异，实现H5、小程序、App的自动化统一

一次定义

只需输入自然语言任务描述，无需针对不同平台编写多套脚本，大幅提升开发效率

处处执行

同一任务可在移动端、PC端、不同操作系统间无缝迁移，真正实现跨终端通用执行

语义抗变化&降低维护成本

界面变更不影响语义理解，适配时间从2小时缩短至10分钟内，显著降低维护成本

当前不足与未来演进方向

05

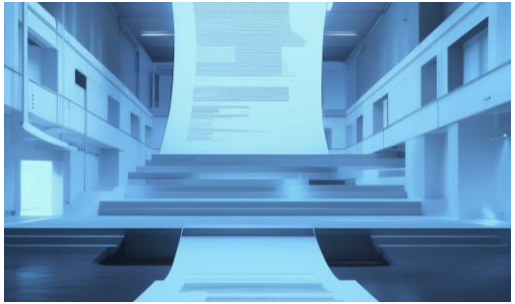
面临延迟较高、小字体识别不准、上下文遗忘等技术瓶颈

推理延迟高

多模态模型平均响应3~5秒，影响高频交互体验，制约实时性要求高的场景落地

小字识别难

OCR对模糊、小字号或变形字体存在漏检，导致关键信息缺失，影响任务准确性



记忆窗口限

上下文长度受限（约32k tokens），超长流程可能出现历史动作遗忘，影响连贯决策

推进端侧轻量化部署，目标将推理延迟降至1秒以内

飞猪旅行

延迟瓶颈

当前云端多模态模型推理耗时3~5秒，影响高频交互体验

端侧部署

探索蒸馏版UI-Tars与TinyLLM本地化运行，减少网络依赖

算力优化

结合设备硬件加速（如无影），提升端侧推理效率与响应速度

目标延迟

实现端到端决策延迟低于1秒，满足实时操作需求

构建强化学习闭环，利用反馈数据持续优化模型推理决策质量

极客邦科技 2026 年会议规划

促进软件开发及相关领域知识与创新的传播

参会咨询

查看会议

北京 · 1200人

QCon 全球软件开发大会

会议时间：4月16-18日

- Agentic Engineering
- AgentOps
- 下一代模型架构与推理优化
- AI 原生基础设施
- 知识工程实践
- AI 安全

深圳 · 1000人

AiCon 全球人工智能开发与应用大会

会议时间：8月21-22日

- Agentic AI
- 轻量化与高效推理
- 多模态应用
- AI + IoT 场景实践
- AI 工业化落地

北京 · 1000人

AiCon 全球人工智能开发与应用大会

会议时间：12月18-19日

- 大模型架构创新
- 多模态 AI 产业融合
- 具身智能
- AI for Science
- 大模型安全

4月 — 6月 — 8月 — 10月 — 12月

上海 · 1000人

AiCon 全球人工智能开发与应用大会

会议时间：6月26-27日

- AI Infra 系统工程
- 多 Agent 协作与实践
- 多模态融合
- 模型训练与推理创新
- 数据平台与特征服务

上海 · 1200人

QCon 全球软件开发大会

会议时间：10月22-24日

- AI Agent
- Vibe Coding
- 智能可观测
- 推理基建
- 模型攻防
- AI x 创造力

InfoQ 极客传媒

飞猪旅行

THANKS