

QCon
全球软件开发大会

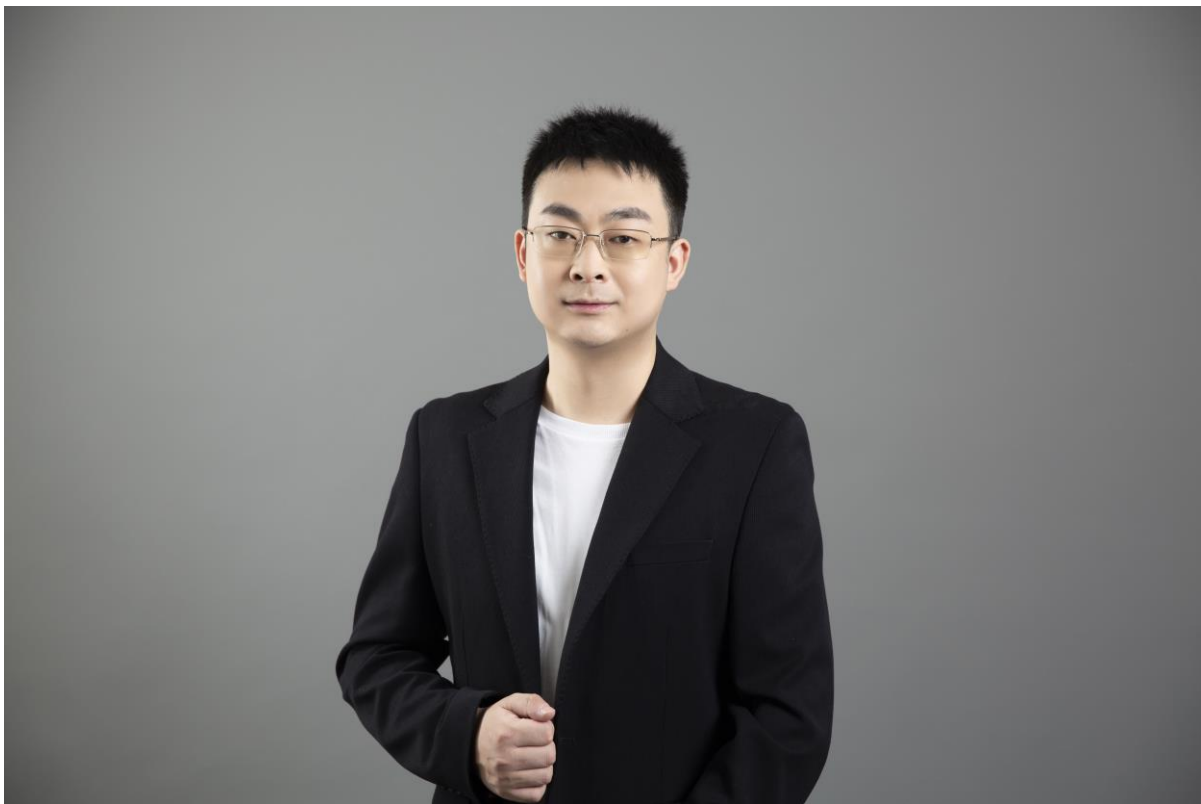
趣丸运维 AI Agent

从 1.0 到 2.0 的自学习进化实战

黄金

基础架构负责人





黄金

趣丸网络-基础架构负责人

现任趣丸科技基础架构组负责人，负责多云基础设施建设和研发效能相关平台。10年工作经验，专注于多云基础设施和AIGC赋能项目。在容器技术、高可用系统架构设计以及人工智能领域积累了丰富的实践经验。积极参与开源贡献，多次在国内知名技术峰会上分享。目前正在积极探索AIGC技术在研发流程中的创新赋能方式。

目录

- 01 机遇与挑战：大模型驱动的智能运维 Agent 新范式
- 02 从 0 到 1：趣丸运维 Agent 1.0 的落地与反思
- 03 自我进化：Agent 2.0 的自学习机制与技术路径
- 04 未来图景：目标驱动的主动式人机协同新方向

📍 北京

QCon

全球软件开发大会

会议时间：4月10-12日

- 大模型赋能 AIOps
- 越挫越勇的大前端
- 云上业务架构演进
- 多模态大模型及应用
- 海外 AI 应用创新实践

📍 北京

AiCon

全球人工智能开发与应用大会

会议时间：6月27-28日

- 端侧智能
- AI Agent
- 多模态大模型
- 金融+大模型

📍 上海

QCon

全球软件开发大会

会议时间：10月23-25日

- AI Agent
- 大模型训练推理
- 端侧 AI
- 搜推深度融合
- 数智金融

4月

5月

6月

8月

10月

12月

📍 上海

AiCon

全球人工智能开发与应用大会

会议时间：5月23-24日

- 通用大模型
- AI Agent
- 垂直领域应用
- 模型可解释性

📍 深圳

AiCon

全球人工智能开发与应用大会

会议时间：8月22-23日

- 模型效率与部署
- 多模态大模型
- 大模型安全
- 智能硬件

📍 北京

AiCon

全球人工智能开发与应用大会

会议时间：12月19-20日

- 通用大模型
- 智能硬件
- LMOps
- 具身智能

极客邦科技 2025 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

01

机遇与挑战

大模型驱动下的智能运维 Agent 新范式

传统智能运维的困境

泛化能力弱，规则系统维护成本高

传统智能运维系统往往针对特定场景设计，缺乏通用性，导致规则系统在面对新问题时需要频繁调整，增加了维护成本和时间。

人机缺乏协同，模型结果无法干预

传统智能运维系统往往缺乏与人工的有效协同机制，模型产生的结果无法得到及时的人工干预和校正，限制了系统的灵活性和准确性。



依赖人工经验，知识传承断层

运维工作严重依赖于个人经验，缺乏有效的知识共享机制，导致经验无法有效传承，新一代运维人员难以快速掌握关键技能。

非结构化数据处理难题

运维中遇到的大量日志和告警信息通常是非结构化的，缺乏有效的工具和技术来高效关联和分析这些数据，影响了问题的快速定位和解决。

LLM Agent 在运维中的机会

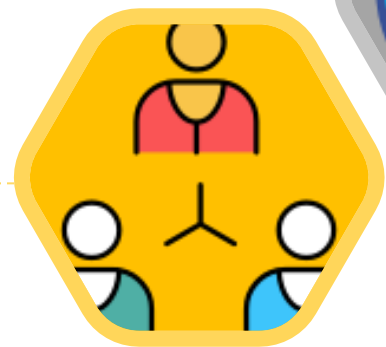
自然语言驱动的运维一体化

利用自然语言处理技术，AI Agent 可以创建一个统一的入口，实现查询、监控和操作的一体化，简化运维流程。



人机协作模式的优化

AI Agent 与人类运维人员的协作模式优化，使得人员可以专注于裁决和把关，而让 Agent 执行繁重的日常运维工作。



认知到决策的智能化闭环

AI Agent 通过从认知到决策的闭环，将传统的自动化升级为智能化，从而提高运维效率和准确性。



面向任务的自适应推理能力

AI Agent 能够根据任务需求进行自适应推理，更稳健地处理各种边界情况和异常，提升问题解决的灵活性。



运维场景的特殊性

01

低容错空间

运维场景中，AI Agent需容错率低以确保系统稳定，执行任务时强制确认操作，过程可控。同时，需配套权限控制、操作审计及快速回滚机制保障系统安全。

03

私域数据为主

AI Agent处理企业私域数据，如术语、流程等，这些数据对企业决策至关重要。它依赖实时内部数据，确保决策准确实用。

02

复杂多模态输入

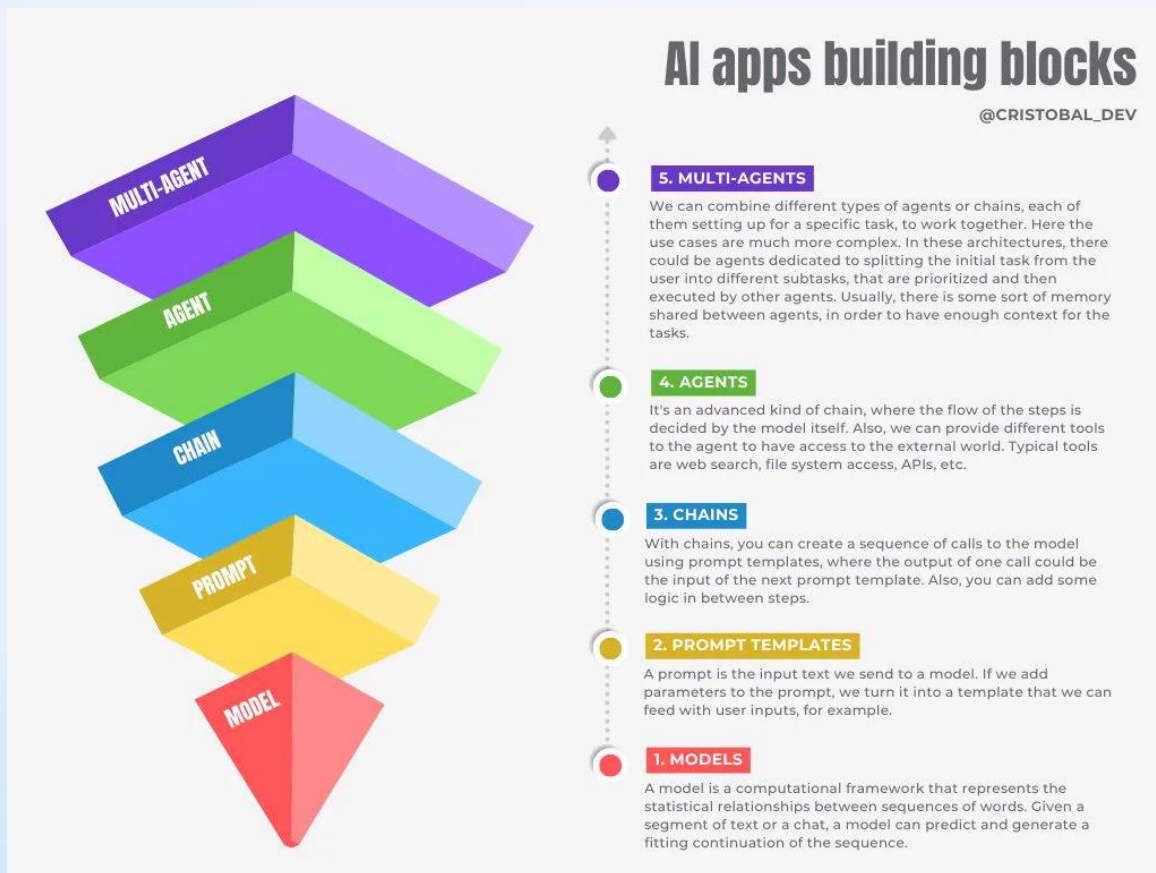
运维AI Agent需处理文本、图片等输入，要有强多模态理解能力，集成算法解析数据以准确决策响应。

04

实时性要求

AI Agent在运维场景需具备实时性，快速响应系统变化，实时监控并处理异常。这要求其有高效数据处理和快速决策能力，以提高运维效率和系统可靠性。

大模型在运维场景的落地思路



五层基石理论

LLM 运维场景落地



- 确定一个步骤
- 测试 LLM 能力
- 自动化步骤
- 解决依赖流程
- 自动化大多数操作
- 闭环场景

大模型在智能运维上的进化方向



支持问答与数据分析

AI Agent通过问答和数据分析提供决策支持，帮助用户理解复杂信息。



协助执行操作

在人类协助下，AI Agent可执行特定任务，减轻工作负担。



主动规划与指导

AI Agent能主动规划并指导人类，完成复杂协作任务。

02 从0到1

智能运维Agent 1.0 落地与反思

AI Agent 技术落地的问题和挑战

稳定性

生成式模型在长对话或复杂任务中可能表现出幻觉、不一致或性能退化。这些问题往往源自模型缺乏因果推理能力、对语境极端敏感以及内部随机性。

成本

设计 AI Agent 时需综合考虑模型大小、推理并行度、硬件选择以及是否采用知识检索或蒸馏等技术，以在效果与成本之间平衡。同时上线文的管理方式、记忆与工具等都会为模型调用带来额外的成本消耗



时效性

运维场景下的决策通常需要实时或准实时响应，若 AI Agent 在信息处理、模型推理或指令下发阶段耗时过长，会直接影响问题定位与恢复效率，导致整体 MTTR 延长。

准确性

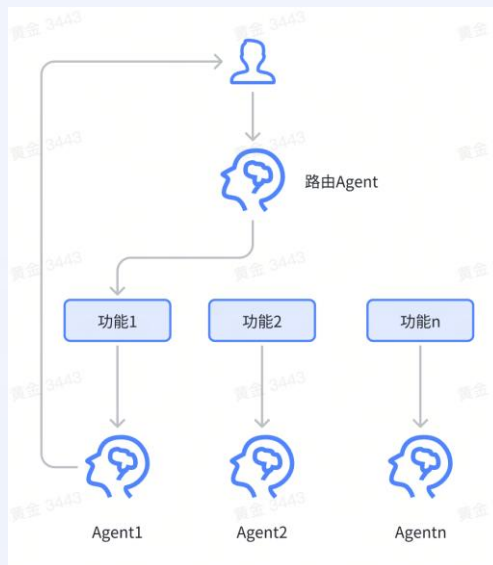
LLM 存在知识截断、知识鸿沟等问题——无法覆盖企业内部专有系统、最新架构与配置数据，易产生错误推断。削弱运维人员信任。

趣丸运维Agent 1.0演进路线



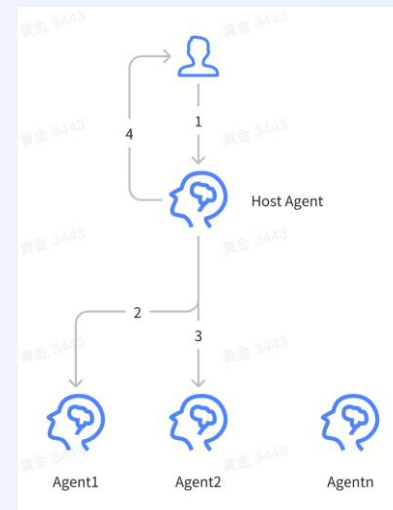
阶段 一

用户决定使用哪个Agent



阶段 二

路由Agent决定使用哪个Agent



阶段 三

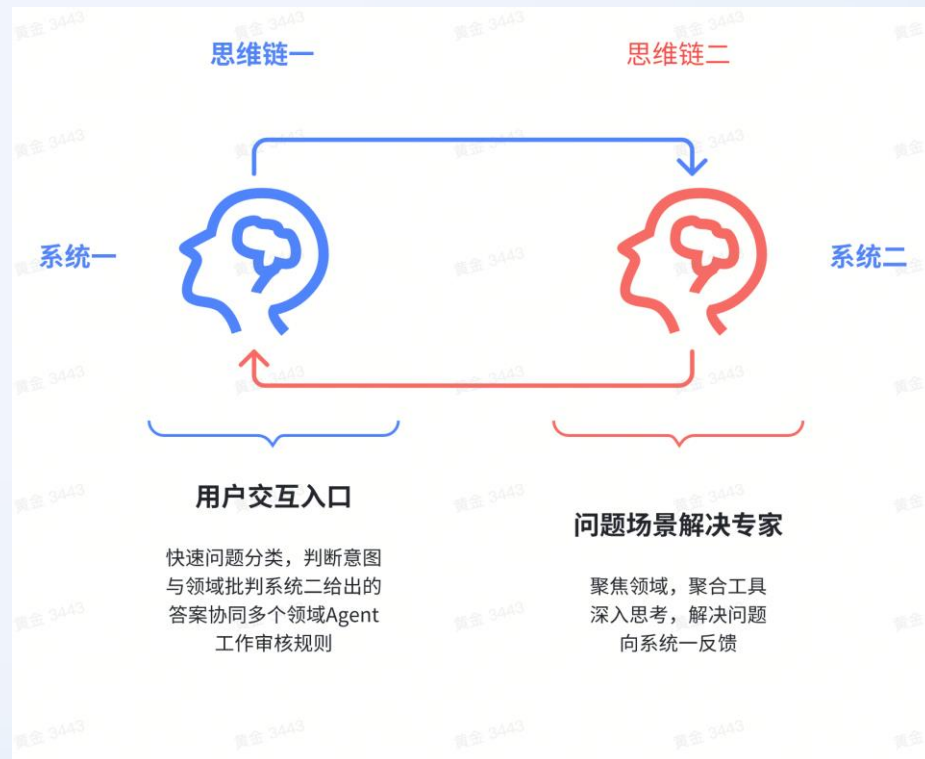
主持Agent协调多个Agent工作

Agent 1.0实现

双思维链模型

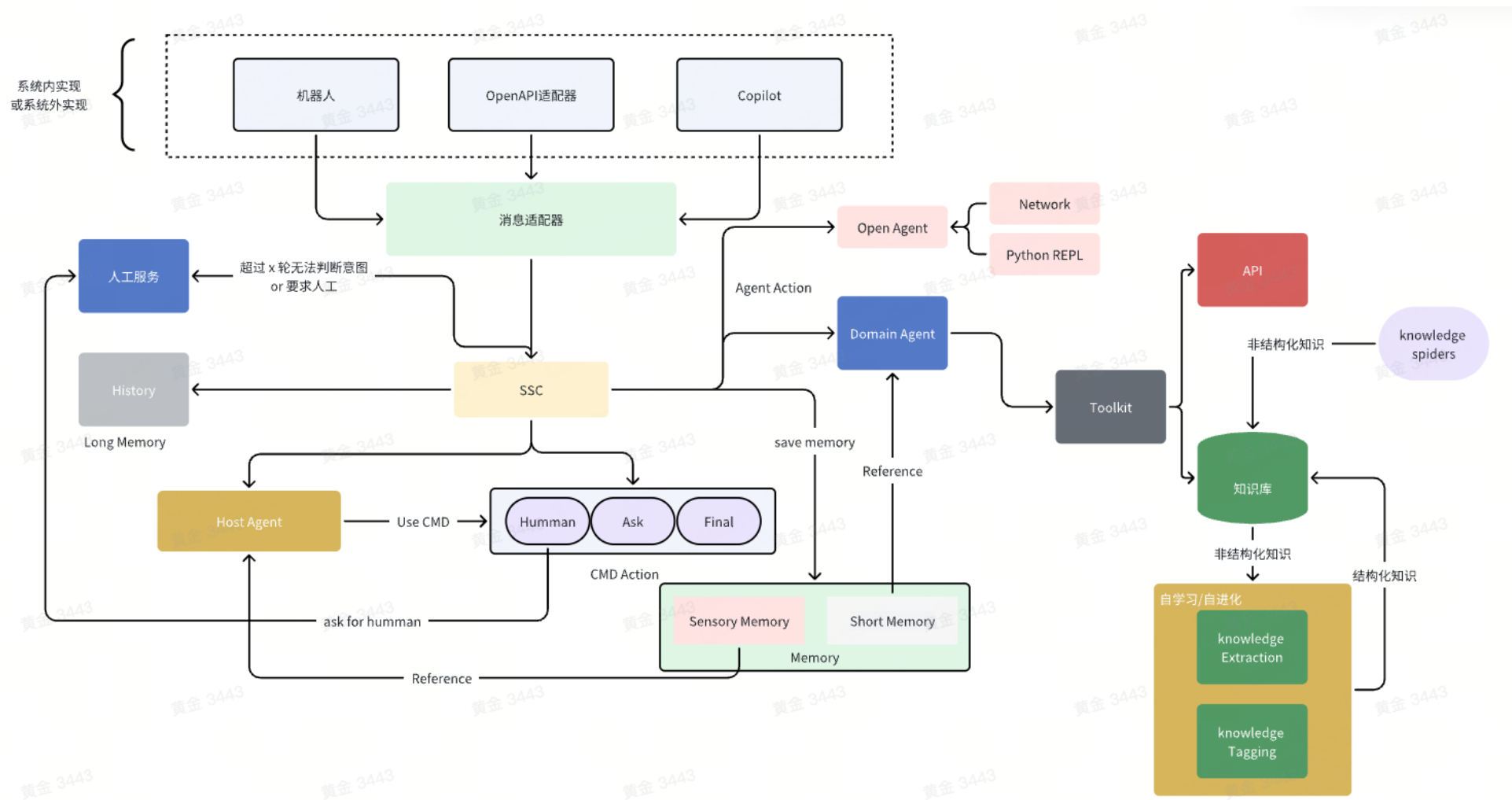


《思考的快与慢》



基于系统一与系统二多Agent协作模型

运维Agent1.0架构



Agent 1.0落地场景与成效



咨询类

- 飞书文档问答
- CMDB查询
- 命令号查询
- 查询监控图表
- 拨测域名
- 应用CPU资源推荐
- 查询事件中心



故障处理类

- 故障根因定位
- 应用日志根因分析
- 应用下游根因分析
- 应用是否变更
- 故障处理提醒与总结
- 非工作时间值守及自助加急电话



资源交付类

- 主机权限申请
- k8s权限申请



其他

- 群聊信息记录与总结



执行

- 删除Pod
- 重启Workload
- 告警指派
- 电话通知
- 抑制告警

智能运维Agent的应用场景-知识问答

12:20

回复 黄金: 域名申请教程

Hi~ Jarvis为您服务

域名申请的步骤如下:

1. 业务方发起域名采购续期
2. 由技术保障部-IT负责采购域名
3. 域名采购后解析权限转技术保障部-基础架构组
4. 技术保障部-基础架构组完成域名ICP备案、公安备案
5. 法务完成网文资质申请

注意: 完成第四步后业务域名机可以使用。根据业务性质

备案信息查询:
公安备案: <http://www.beian.gov.cn/portal/index.do>
ICP备案: <https://beian.miit.gov.cn/#/Integrated/record>

域名使用和维护包括:

- 域名到期续期
- 域名使用合规
- 域名解析权限
- 域名备案
- 域名网文备案
- 域名公安备案
- 域名主体信息调整

个人域名转移到公司:

- 转移到指定主体, 发起域名信息变更
- 更新域名备案信息
- 更新公安备案信息
- 更新网文备案信息
- 域名解析权限调整公司dns

详细请查看文档: <https://q9jvw0u5f5.feishu.cn/wiki/wikcnUBOrbDkbXTSC7PZBG35R6f>,
<https://q9jvw0u5f5.feishu.cn/wiki/wikcnvkmmDzxLRxbKN4WrCclroe>,

域名申请教程

Jarvis

你太啰嗦了

Jarvis

1 条回复

12:21

回复 黄金: 你太啰嗦了

Hi~ Jarvis为您服务

对不起, 我会尽量简化回答。关于域名申请, 主要步骤如下:

1. 业务方发起域名采购续期
2. IT部门负责采购域名
3. 域名采购后, 基础架构组负责域名解析和备案
4. 法务部门完成网文资质申请

具体操作和细节可以参考相关文档。

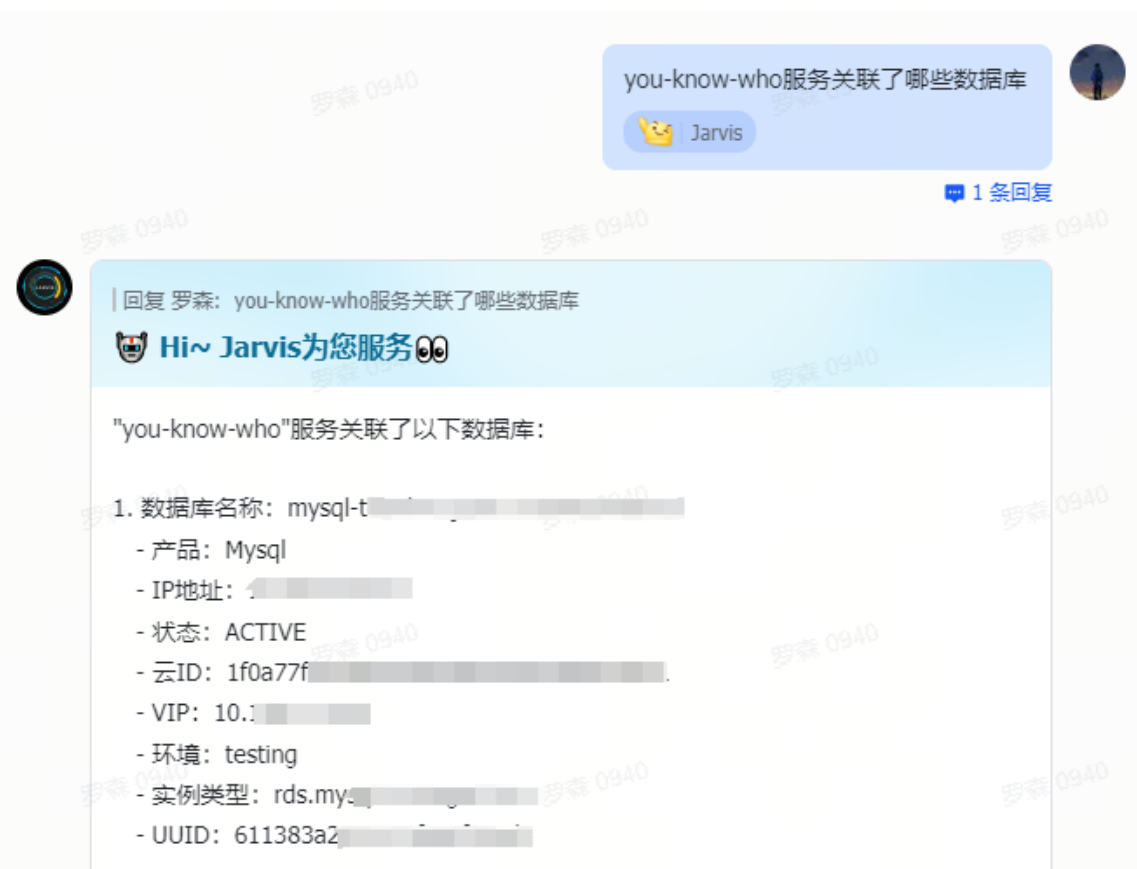
注: 本回答来源于知了

有帮助

无帮助

更换角色

智能运维Agent的应用场景-资源查询



智能运维Agent的应用场景-异常定位

诊断报告 - muse-social-community-auditsub-tt-dev-jupiter-f5l

发生事件

Type	Reason	Age
Warning	BackOff	1小时前

发生的错误

back-off 5m0s restarting failed container=service pod=muse-social-community-auditsub-tt-dev-jupiter-f5l

处理建议

1. 检查Pod中的容器名称为muse-social-community-auditsub-tt-dev-jupiter-f5l
2. 确保容器的执行文件格式正确。
3. 如果格式错误，请修复执行文件格式。
4. 重新启动Pod以应用更改。
5. 检查容器的退出代码是否为1。
6. 如果退出代码为1，请根据错误日志查找问题的根本原因。
7. 根据问题的根本原因提供相应的解决方案。

崩溃日志

```
1 exec /muse-social-community-auditsub: exec for
2
```



注: 本回答来源于知了，如回答错误请点击下方按钮发送意见反馈或转人工服务

错误反馈

venjie-54f48576b4-tnqhz怎么了

1条回复

人员: [REDACTED]

Success

svr该应用自身异常，请尽快排查。

需要进一步分析。

常:

GroupMemberList ret:-8

需要进一步分析。

ild fail. ret = -10014, game id = 15

具体原因需要进一步分析。可能

组成员失败、获取游戏公会信息失败
网络连接异常。需要及时排查故障

智能运维Agent的应用场景-辅助提单

王恒 12月12日 14:02

申请主机[REDACTED]权限，账号godman

Jarvis

Jarvis 12月12日 14:02

| 回复 王恒: 申请主机[REDACTED]7权限，账号god...

Hi~ Jarvis为您服务🐼

- 提交结果: 主机权限申请工单提交成功
- 审批状态: 已通知主机业务运维审批

📌注: 本回答来源于知了

👍有帮助 🙏无帮助

申请K8S权限

知了 - Jarvis

1 条回复

| 回复 王恒: 申请K8S权限

Hi~ Jarvis为您服务🐼

请提供以下信息以便我帮助您申请K8S权限:

1. 您需要申请权限的kubernetes集群名称是什么?
2. 您需要申请权限的命名空间是什么?
3. 您需要申请的角色名称是什么? (角色名称必须为view、edit两种角色之一)

📌注: 本回答来源于知了, 如回答错误请点击下方按钮发送意见反馈或转人工服务

✖ 错误反馈

集群k8s-tc-bj-1-yunwei

知了 - Jarvis

1 条回复

| 回复 王恒: 集群k8s-tc-bj-1-yunwei

Hi~ Jarvis为您服务🐼

请提供以下信息以便我帮助您申请K8S权限:

1. 您需要申请权限的命名空间是什么?
2. 您需要申请的角色名称是什么? (角色名称必须为view、edit两种角色之一)

📌注: 本回答来源于知了, 如回答错误请点击下方按钮发送意见反馈或转人工服务

✖ 错误反馈

命名空间 infra 角色view

知了 - Jarvis

1 条回复

| 回复 王恒: 命名空间 infra 角色view

Hi~ Jarvis为您服务🐼

用户王恒已成功申请k8s-tc-bj-1-yunwei集群下命名空间infra的view角色权限, 工单已提交, 正在等待管理员审批。

📌注: 本回答来源于知了, 如回答错误请点击下方按钮发送意见反馈或转人工服务

✖ 错误反馈

智能运维Agent的应用场景-故障管理

知了 - Jarvis 机器人 | 百事通

@所有人 🚨 故障已持续10min, 请各位加急处理!

知了 - Jarvis 机器人 | 百事通 昨天 20:28

故障信息及进展同步

故障进度概述

- 问题描述:
- 影响的功能: 小游戏
- 故障开始时间:
- 当前处理进度:

故障中止 故障结束

@Jarvis 故障结束

Jarvis 机器人 | 百事通

收到, 故障结束!

【故障】2023年12月12日 TT语音话筒下方出现空白头像故障

故障开始时间: 2023-12-12 10:44:02

故障定位时间: 2023-12-12 11:24:13

故障恢复时间: 2023-12-12 12:17:38

故障处理时长: 93min

故障原因: 问题的原因是'踢下麦'操作中, uid为0导致麦下列表异常及空白头像的出现。

故障影响: 这个故障影响了TT语音的语音聊天界面的正常显示, 可能会降低用户的语音聊天体验。

12:18

故障分析:

- 采取措施及时性:
- >措施1: 在发现故障后, 队伍启动了错误复现和定位措施, 这是非常及时的。
- >措施2: 队伍在理解故障原因后, 主动设定了针对uid为0的后端过滤措施, 并在安卓端实行了此操作, 这是非常有效的。
- 协作流程: 团队成员之间的协作流畅, 开发与运维团队能够快速联手解决问题。
- 展现出的问题: 在故障发生后, 团队需要改进错误预测系统, 预防此类问题在未来再次发生。

故障报告

Agent 1.0落地问题与反思

私域术语理解不足问题

在处理企业私域的特定术语和流程时，Agent 1.0表现出理解不足，这直接影响了其回答问题的准确性和满足企业需求的能力。

长期任务规划的挑战

Agent 1.0在执行长期任务时，由于缺乏有效的规划，难以保持跨步骤的一致性，导致任务执行效率和效果不佳

多 Agent 协作信息丢失

在多 Agent 协作的场景中，信息丢失问题凸显，关键信息未能得到充分记忆和利用，影响了协作效率和成果质量。

人类介入与反馈机制缺失

Agent 1.0在运行过程中，由于缺少人类的高质量反馈和裁决，难以及时修正错误和优化性能，导致问题持续存在。



问题

Agent 1.0落地问题与反思

三大问题

私域数据

- 听不懂内部俗语
- 不了解组织约定
- 不遵循流程、规范和内部经验

上下文

- 缺乏人员上线文，数据范围、任务参数错误
- 多会话依赖无记忆，需要人员反复输出参数
- 多Agent协作，参数不全，任务使用默认参数

运维协作

- 通过聊天方式授权/校验，完全依赖LLM，过程不可靠
- 运维难以参与任务执行过程，反馈可能导致上下文混乱



1

私域知识

通用LLM，缺乏对企业内部知识的了解



2

上下文

人员背景、任务背景、内部协作上下文决定了任务执行的准确度



3

运维协作

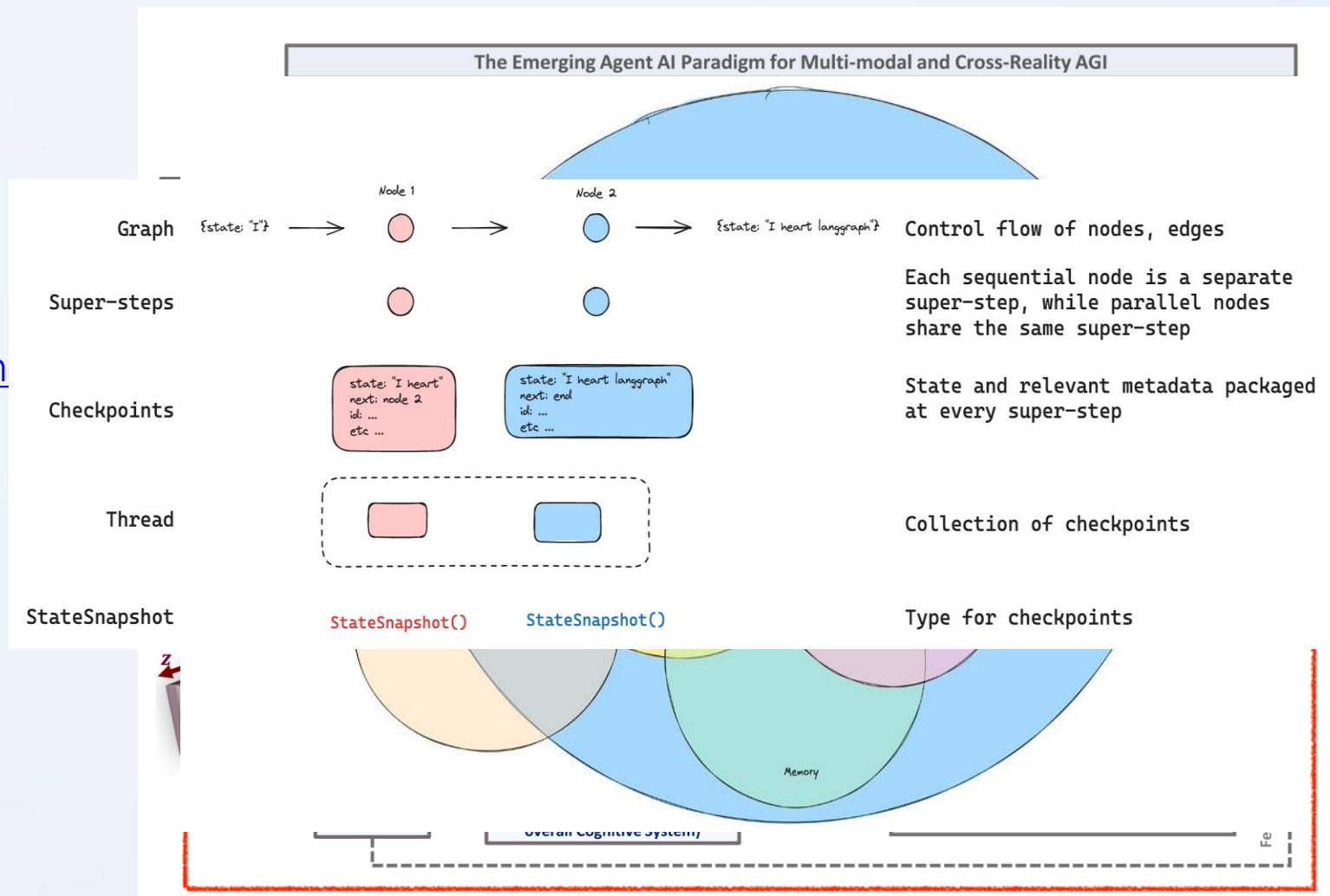
安全、可靠的协作机制是代替运维执行操作的关键保障，用户反馈是决定任务能否继续的关键。

03 自我进化

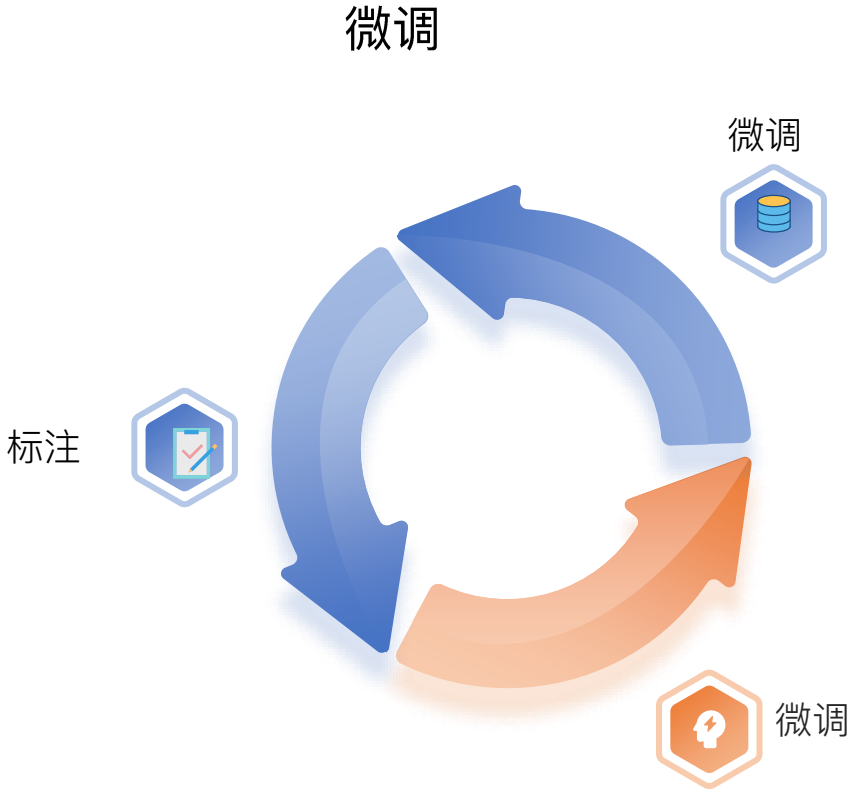
运维 Agent 2.0 的自学习机制与技术路径

新技术带来的新思路

- 一. [AI Agent综述](#)
- 二. [Everything is context engineering](#)
- 三. [checkpointing](#)



私有数据问题



工具调用能力

如何有效地使用工具?

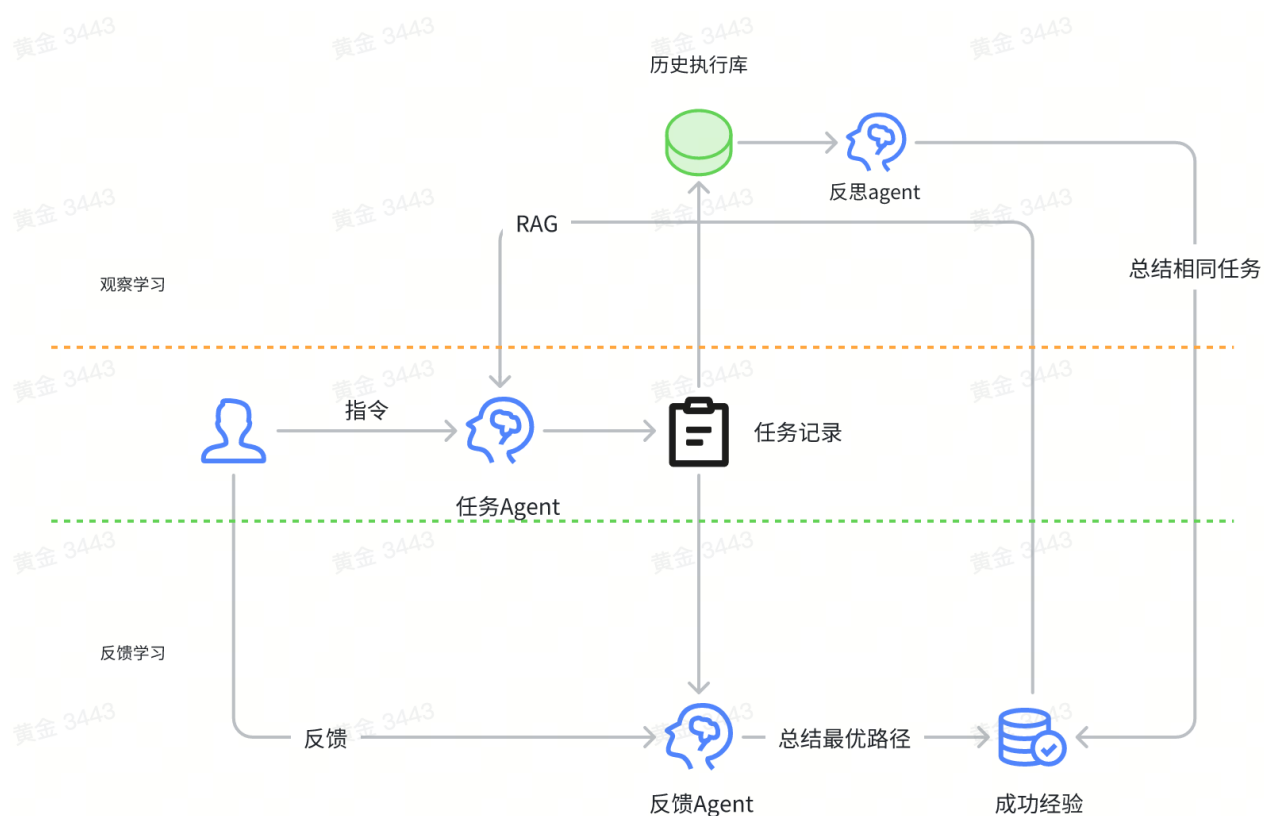
参数配置
配置工具参数以优化性能并满足特定需求。



DeekSeek R1带来的启示:

使用少量高质量监督数据作为冷启动，再结合强化学习微调，能够显著提升大型语言模型的推理性能

自学习流程



反馈->数据->效果

```
{
  "steps": [
    {
      "id": "int, 计划步骤的唯一标识符, 每次都是从1开始递增",
      "plan": "string, 步骤的详细描述, 包含所需参数信息, 包含当前步骤要实现的目的, 不包含评价性内容",
      "action": "string, 工具调用格式, 使用Python函数调用语法, 如: action(arg1=number, arg2='escaped string value', arg3=None), 可选参数可以不填充"
    }
  ]
}
```

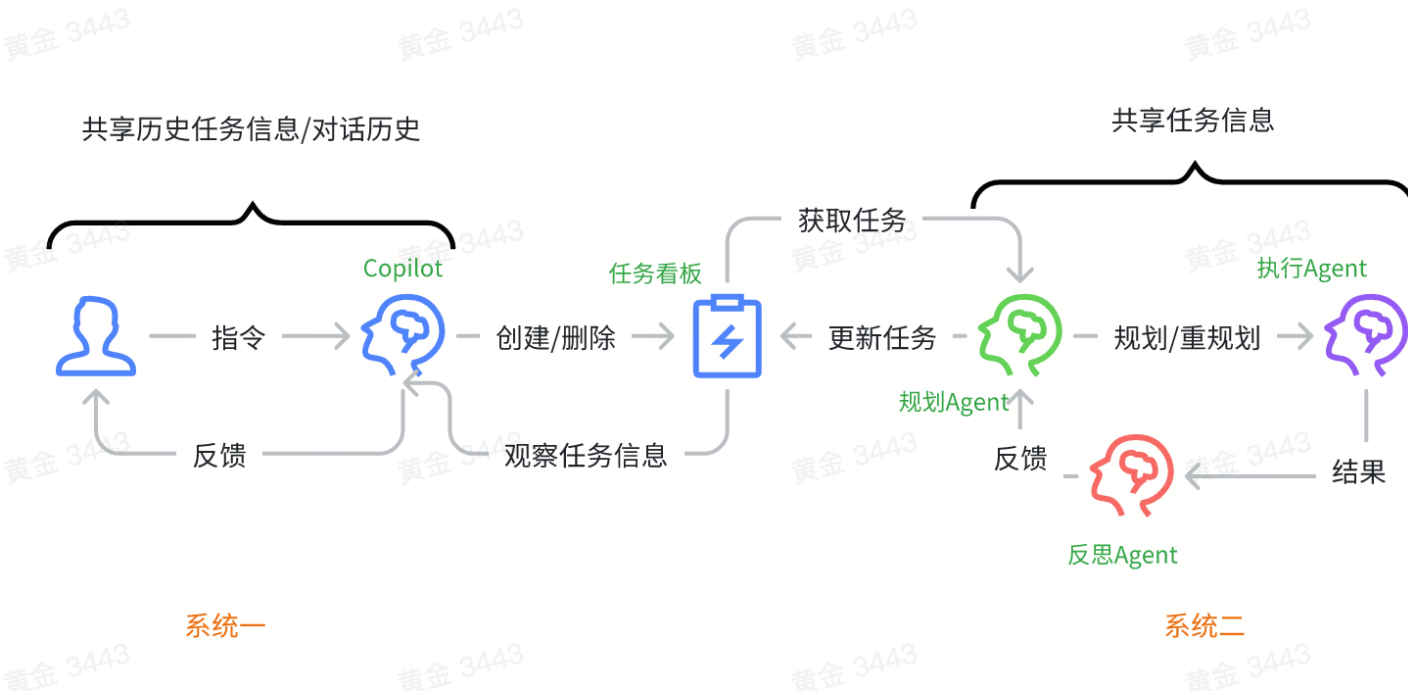
一种格式 · 三种能力 (GRPO)

GRPO (Generalized Reinforced Policy Optimization)

基于反馈强化的自学习框架, 持续优化 Agent 的执行与推理。

1. Function Call: 学习工具调用与参数化。
2. 基础知识: 积累与优化领域经验。
3. 推理规划: 任务分解与策略演化。

上下文工程



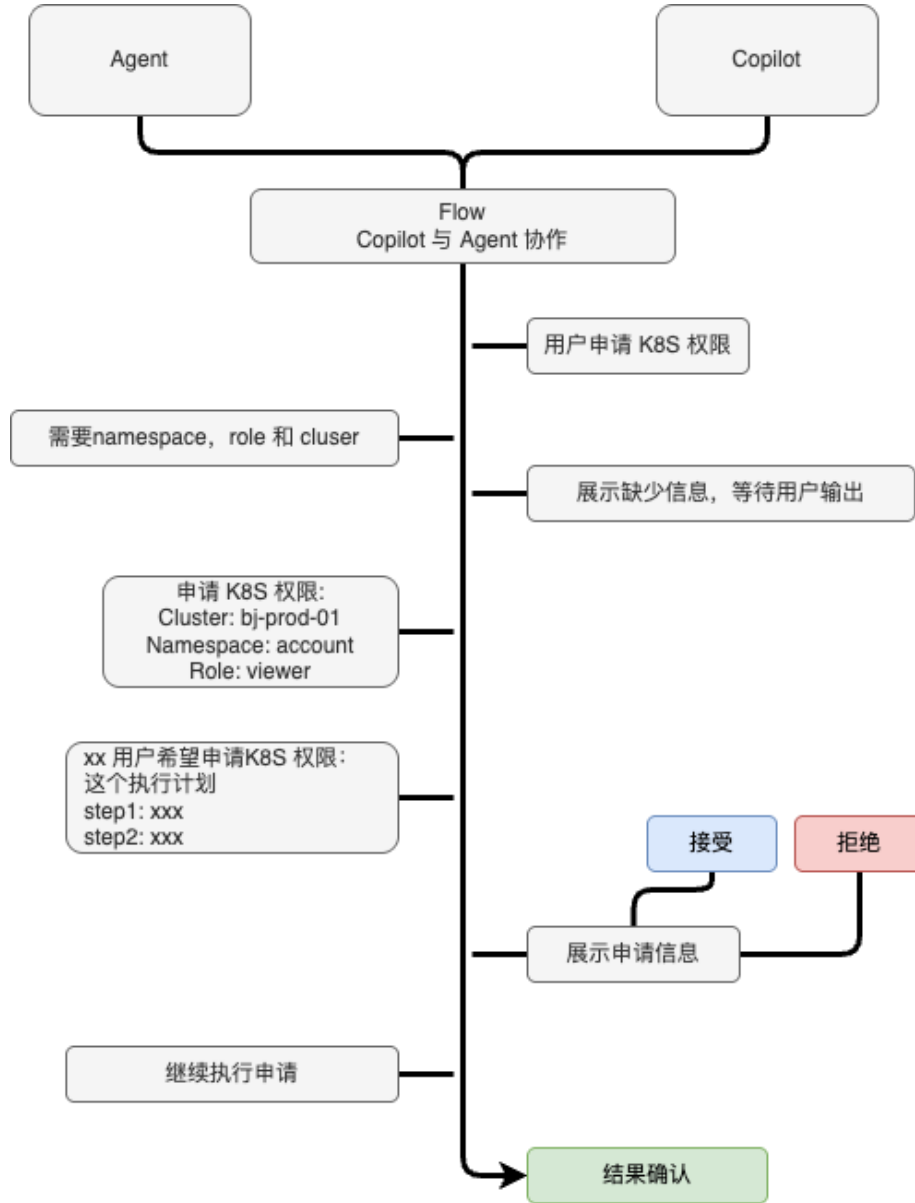
对话与任务隔离



历史压缩

- 同一意图共享完成上下文
- 不同意图压缩上下文

上下文工程-上下文隔离



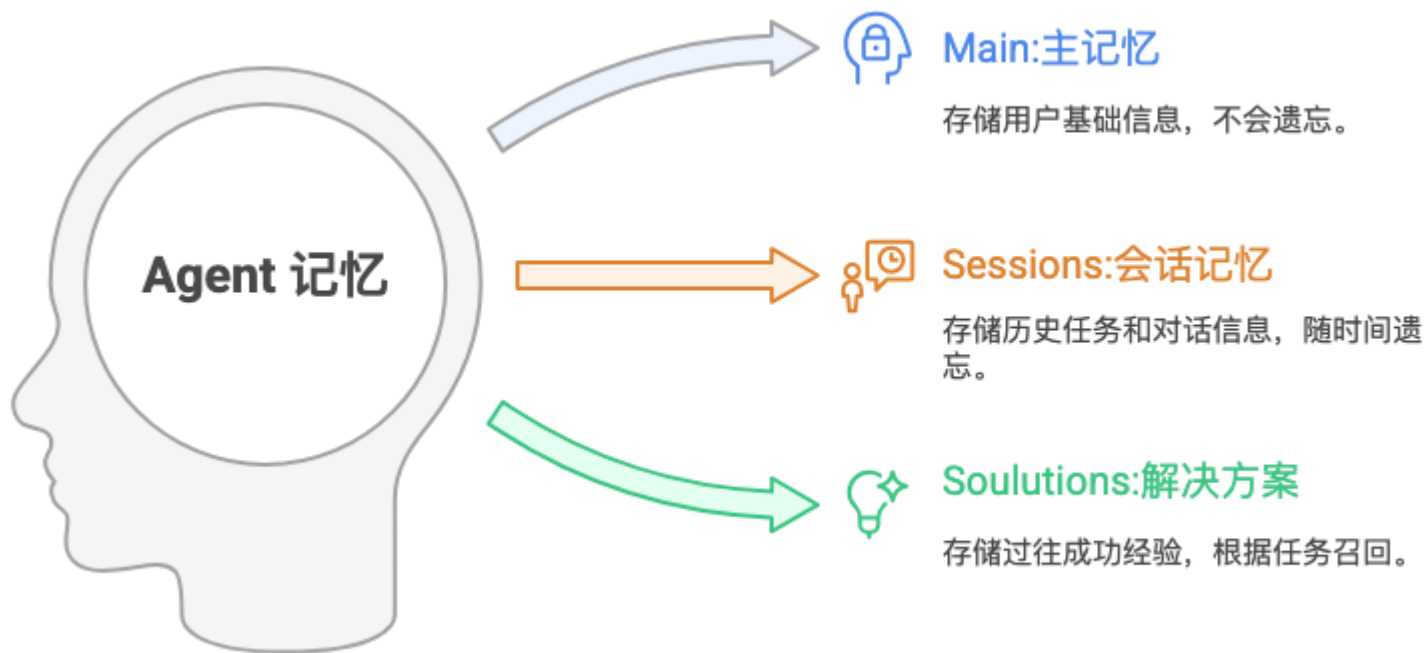
Copilot: 与用户持续对话, 理解用户意图, 引导用户推进任务完成
Agent: 规划任务, 分析任务必要条件, 执行具体步骤

Copilot与用户共享对话Context
Copilot与Agent共享任务Context
任务Context通过任务面板管理

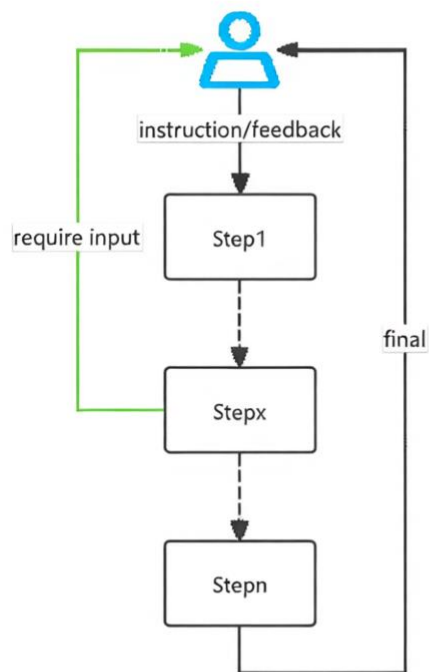
上下文工程-记忆

存储记忆是为了召回，召回方式部分决定记忆存储方式。

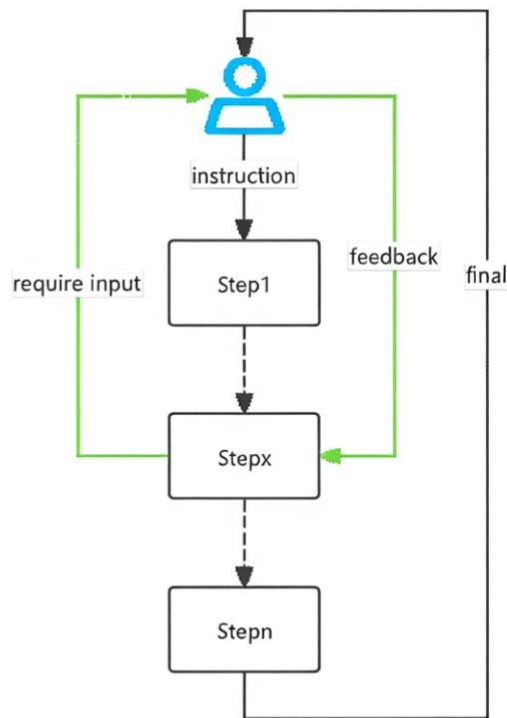
- 必须召回
- 可选召回



人机协作



LLM Return

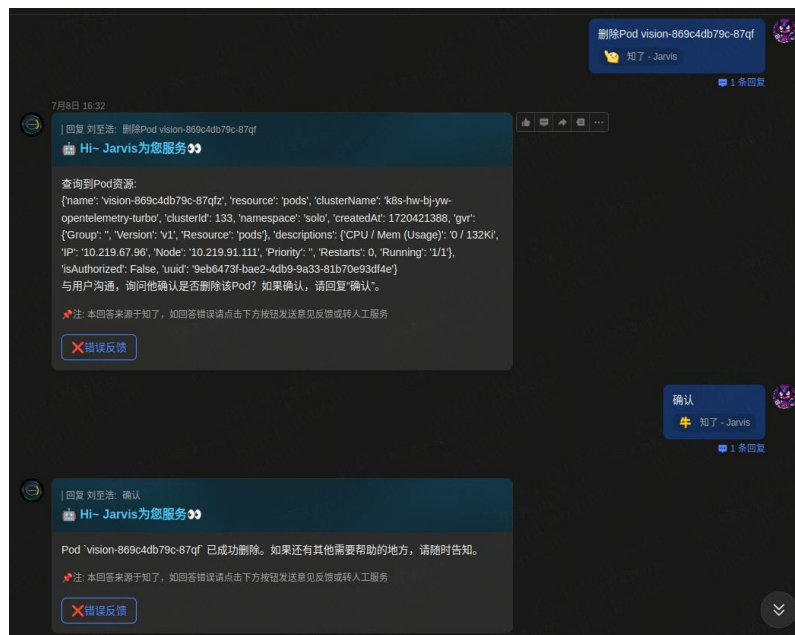


Human In The Loop

四种设计模式实现方式：

- 批准或拒绝：关键步骤前暂停审查，可阻止执行并采取其他操作，通常涉及人工输入路由。
- 编辑状态：审查编辑状态，用于纠错或更新，通常涉及人工输入。
- 审查工具调用：工具执行前暂停审查编辑工具调用。
- 验证人工输入：继续前暂停验证人工输入。

人机协作



LLM Return方式

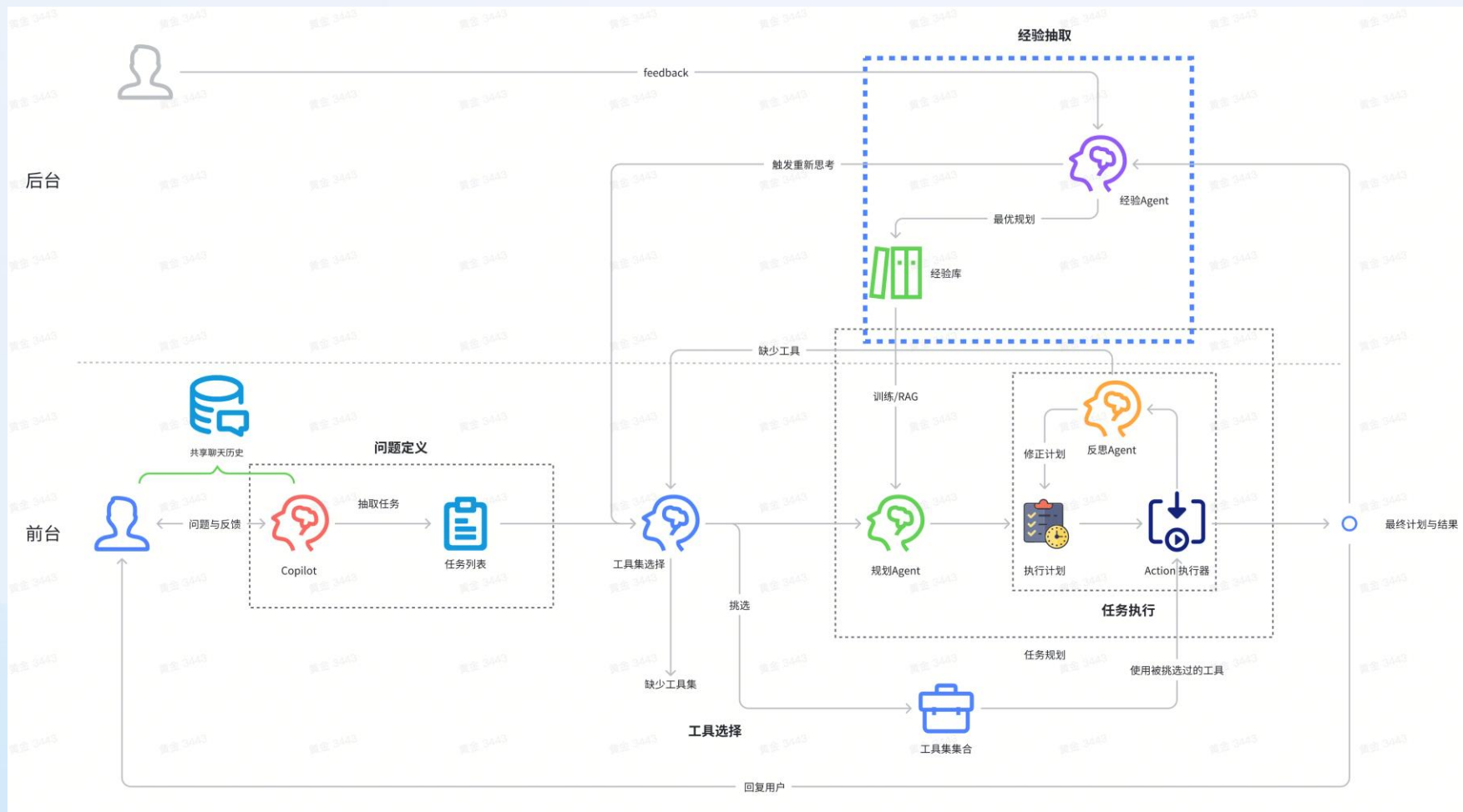
依赖LLM能力与Prompt
易出现出现跳过、反复确认等问题



HITL方式

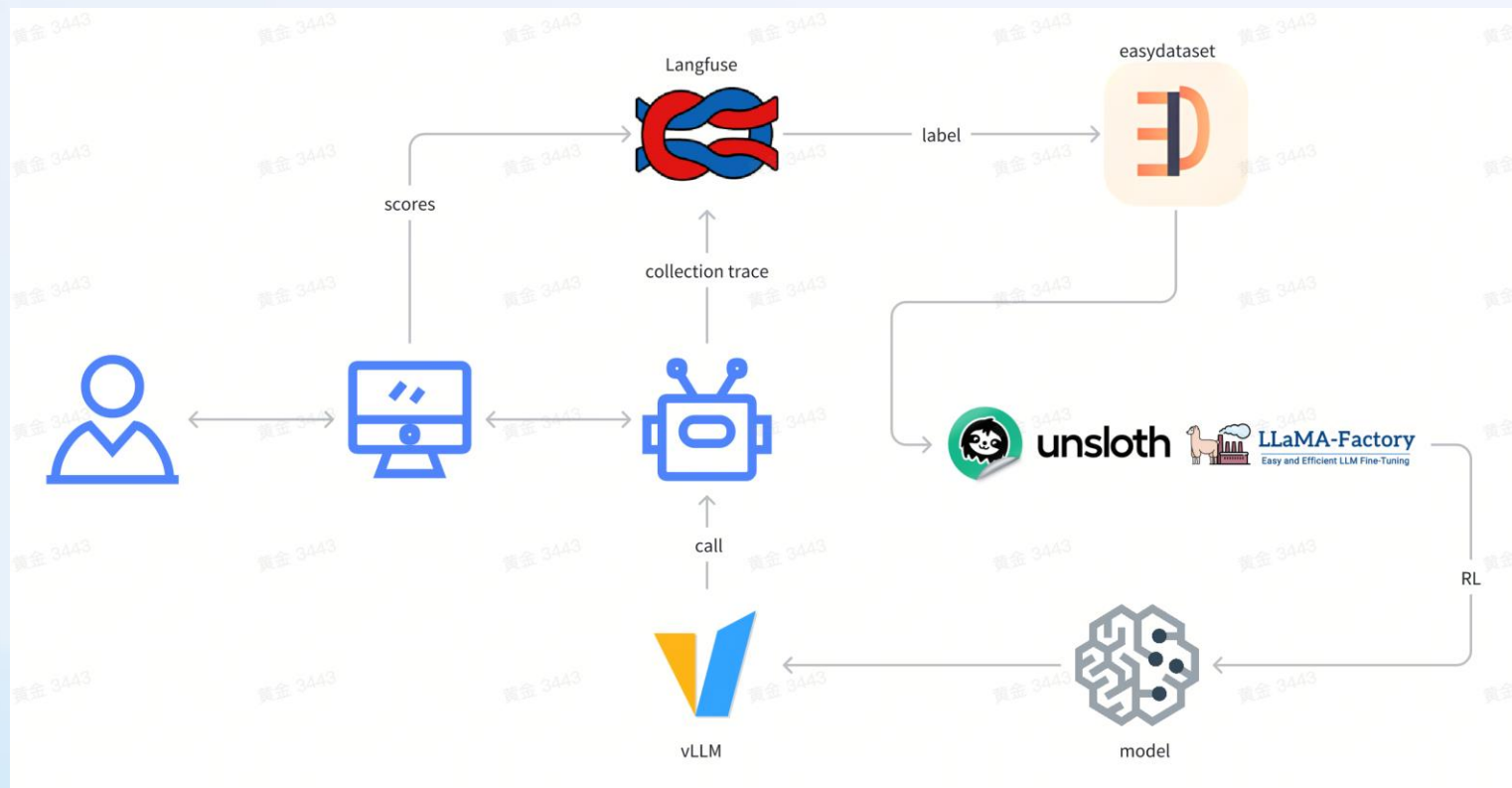
不依赖LLM能力, 完全可控

Agent2.0 架构



高质量的人类标注数据对模型效果提升具有基础性作用——它是模型学到“正确规律”、展示“强泛化能力”、避免“偏差 / 失效情况”的关键

Agent数据循环的构建



标注:

- 自动化标注
通过大量未反馈错误的例子，模型自己总结标注
- 用户反馈标注
通过用户主动反馈标注数据

构建从用户反馈到效果优化的闭环数据链路是确保Agent落地效果的重要一环

Agent数据循环的构建

公司域名申请流程是什么样的

Jarvis

completed

- 用户huangjin@52tt.com想要查询公司域名申请流程的具体步骤和要求。
- mix_query(query='公司域名申请流程')
- FINAL(input='\$1.output')

公司域名申请流程如下:

- 进入【域名申请流程系统】**
 - 操作路径: 通过内部系统提交域名申请, 需提供项目名称、所属业务线、所属项目组、业务环境、项目用途、域名、解析IP、申请人、项目负责人、期望交付时间等必要信息。
- 提交【申请】**
 - 需填写内容: 包括域名、用途、解析IP、申请人、项目负责人等字段, 同时需邮件发送给leader审批和DNS管理员配置解析, 抄送运维邮件组。
- 等待审批**
 - 审批流程包括法务负责人何亮的审批, 基础架构组的审批 (许定琳/朱少华), 以及安全审批人的补充检查。
- 确认结果**
 - 验证方式: 域名配置完成后, 通过DNS解析验证是否生效, 并确认域名是否备案。

参考文档:

- 域名申请/变更/回收<https://q9jvw0u5f5.feishu.cn/wiki/wikcnxQlsUtR0ptQovNzFjxKvFe>

👍 赞

错误原因

用户反馈界面

Node display

Timeline

planner ID

2025-09-28 20:57:57.685

Latency: 9.83s Env: production

Preview

```
plans: [
  {
    message_id: ""
    chat_type: "web"
  }
]
```

Output

```
{
  plans: [
    {
      status: "step"
      content: "计划已生成, 开始执行"
      reason: ""
      steps: [
        {
          id: 1
          content: "使用融合查询工具查询调用链系统使用方法"
          action: "mix_query(query='如何使用公司的调用链系统')\"
          result: null
        }
      ]
      doing_step: {
        id: 1
        content: "使用融合查询工具查询调用链系统使用方法"
        action: "mix_query(query='如何使用公司的调用链系统')\"
        result: null
      }
      past_steps: [
      ]
      result: ""
    }
  ]
}
```

过程数据采集

在构建 Agent 数据循环时, 不能只记录结果, 更要完整采集模型的推理与决策链路, 才能实现可追溯的问题分析与持续的策略优化。

三大关键指标



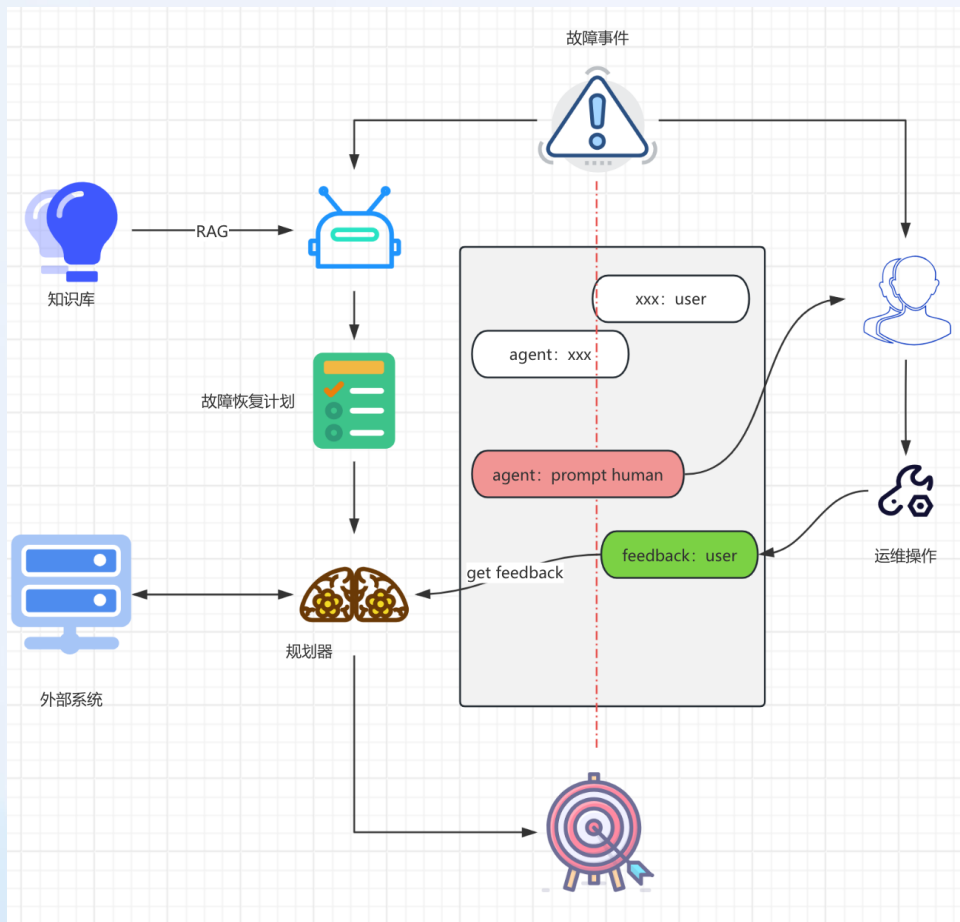
视频演示-WEB端展示



04 未来图景

目标驱动的主动式人机协同新方向

目标驱动



主动故障恢复协作流程

基于共同目标的主动思考

› 基于共同目标/利益

- 角色预设

› 基于个人价值

- 个人信息

主动思考

prompt 人类，从人类得到反馈与协助



主动故障助手

- ✓ 主动识别/执行问题
- ✓ 提出建议、主动推进故障进度



分险治理私人助手（基于个人目标/数据）

- ✓ 定时巡检风险
- ✓ 推荐解决方案
- ✓ 一站式解决

📍 北京

QCon

全球软件开发大会

会议时间：4月10-12日

- 大模型赋能 AIOps
- 越挫越勇的大前端
- 云上业务架构演进
- 多模态大模型及应用
- 海外 AI 应用创新实践

📍 北京

AiCon

全球人工智能开发与应用大会

会议时间：6月27-28日

- 端侧智能
- AI Agent
- 多模态大模型
- 金融+大模型

📍 上海

QCon

全球软件开发大会

会议时间：10月23-25日

- AI Agent
- 大模型训练推理
- 端侧 AI
- 搜推深度融合
- 数智金融

4月

5月

6月

8月

10月

12月

📍 上海

AiCon

全球人工智能开发与应用大会

会议时间：5月23-24日

- 通用大模型
- AI Agent
- 垂直领域应用
- 模型可解释性

📍 深圳

AiCon

全球人工智能开发与应用大会

会议时间：8月22-23日

- 模型效率与部署
- 多模态大模型
- 大模型安全
- 智能硬件

📍 北京

AiCon

全球人工智能开发与应用大会

会议时间：12月19-20日

- 通用大模型
- 智能硬件
- LMOPs
- 具身智能

极客邦科技 2025 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

THANKS

大模型正在重新定义软件

Large Language Model Is Redefining The Software

