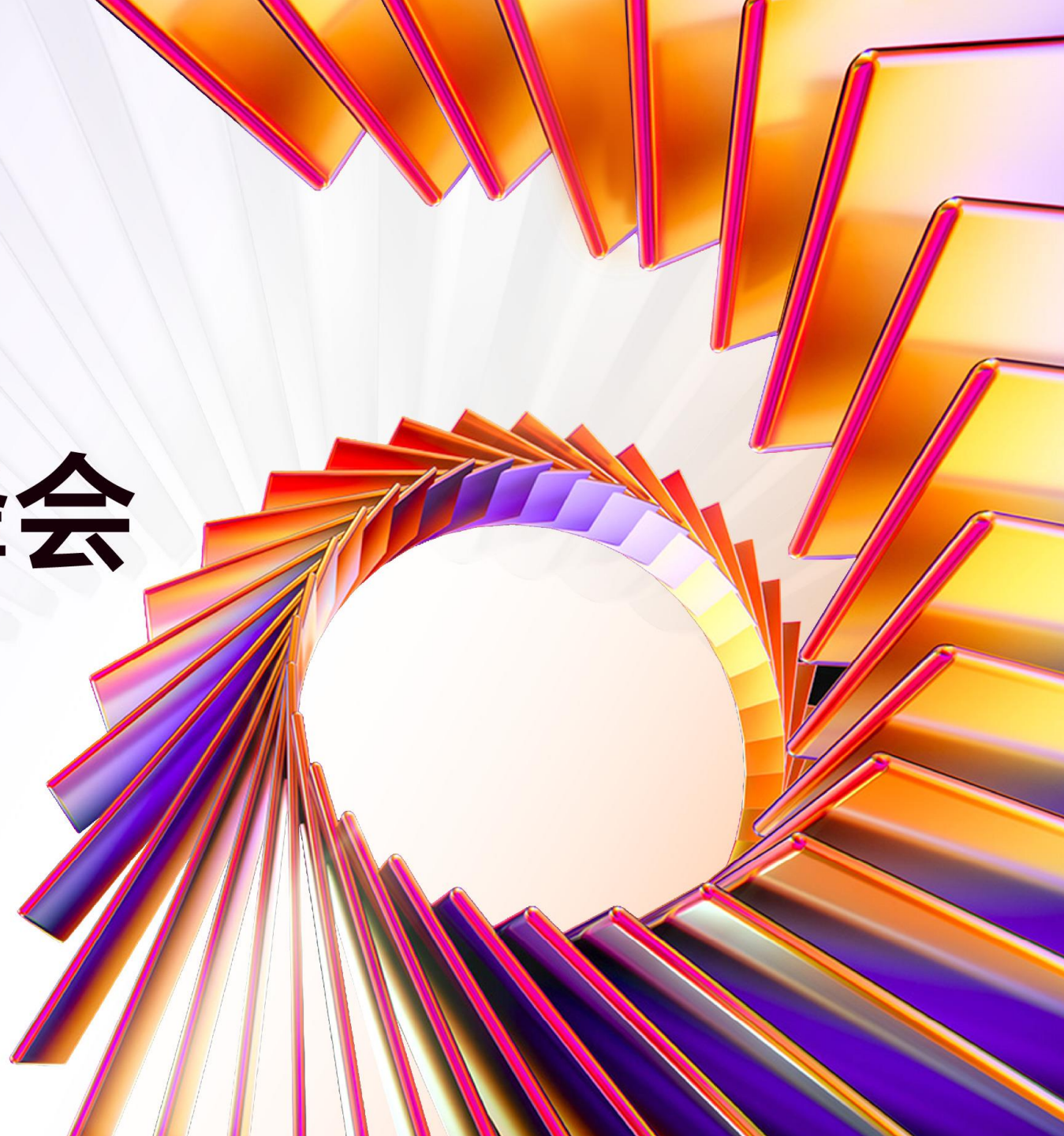
AiDD 6th 2025

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



AiDD 7th 2025

2025 AI+研发数字峰会

拥抱AI 重塑研发

AI+ Development Digital Summit

查看会议详情

下一站预告 08/08-09 | 北京站 11/14-15 | 深圳站

北京站论坛设置

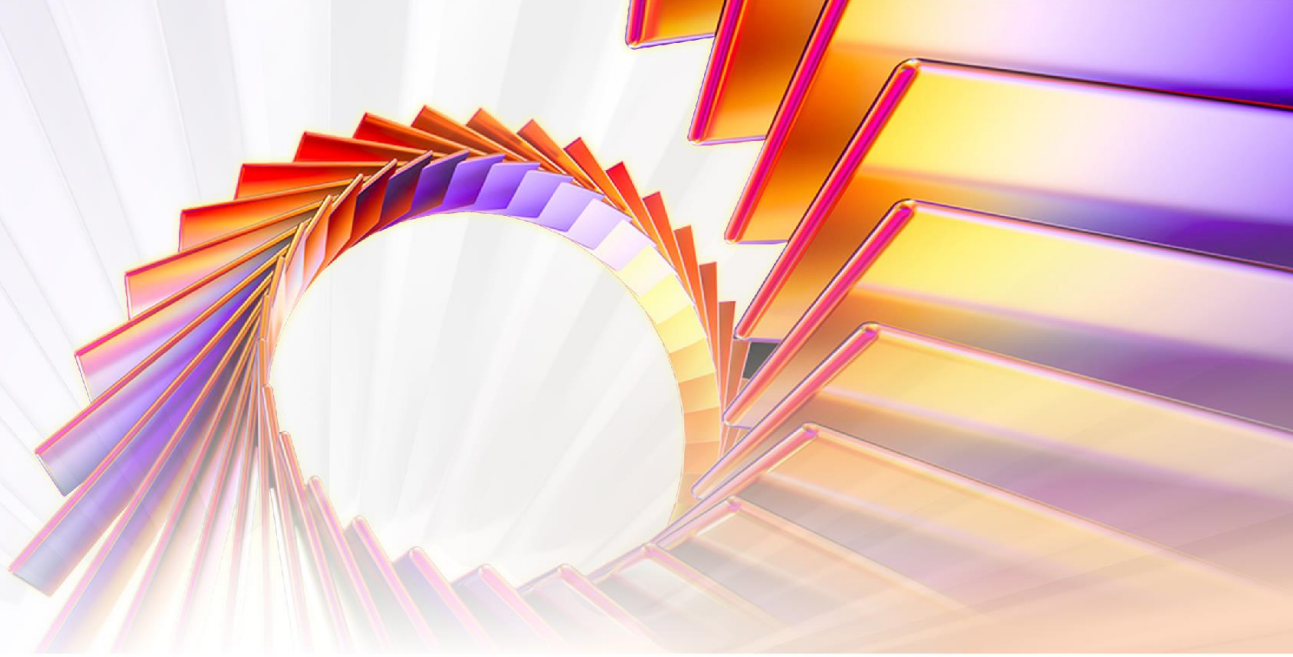
- AI+ 金融业务创新
- 大模型安全与对齐
- 具身智能与机器人
- 大模型和 AI 应用评测
- 智能需求工程
- 大模型应用开发框架与实践
- 代码生成及其改进
- 智能存储与检索技术
- 智能体与研发效率工具
- 智能体经济（Agentic Economy）
- AI+ 新能源汽车
- 下一代知识工程
- AI 产品运营与出海策略
- 智能测试工具的开发与应用
- AI 前沿技术探索与实践

AiDD 6th 2025 | 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 LangGraph 框架和 Ambient智能体构建企业级智能平台

张海立 | LangChain Ambassador

张海立（沧海九粟）

LFAPAC 开源布道师、LangChain Ambassador

《LangChain实战》、《LangGraph实战》作者，LangChain 官方大使，LFAPAC 开源布道师。B 站万粉 UP 主，开源爱好者，长期关注和致力于云原生和前沿互联网技术的技术落地和推广。

曾就职于英特尔亚太研发有限公司，担任高级研发经理和架构师。

目录
CONTENTS

PART 01

Ambient 智能体的核心理念

Ambient 智能体的核心理念

从被动响应到主动感知的范式转变

Ambient 智能体是一种新型的 AI 代理模式，具有以下关键特性：

事件驱动

不依赖用户主动发起对话，而是响应环境中的各种事件和信号

多任务并行

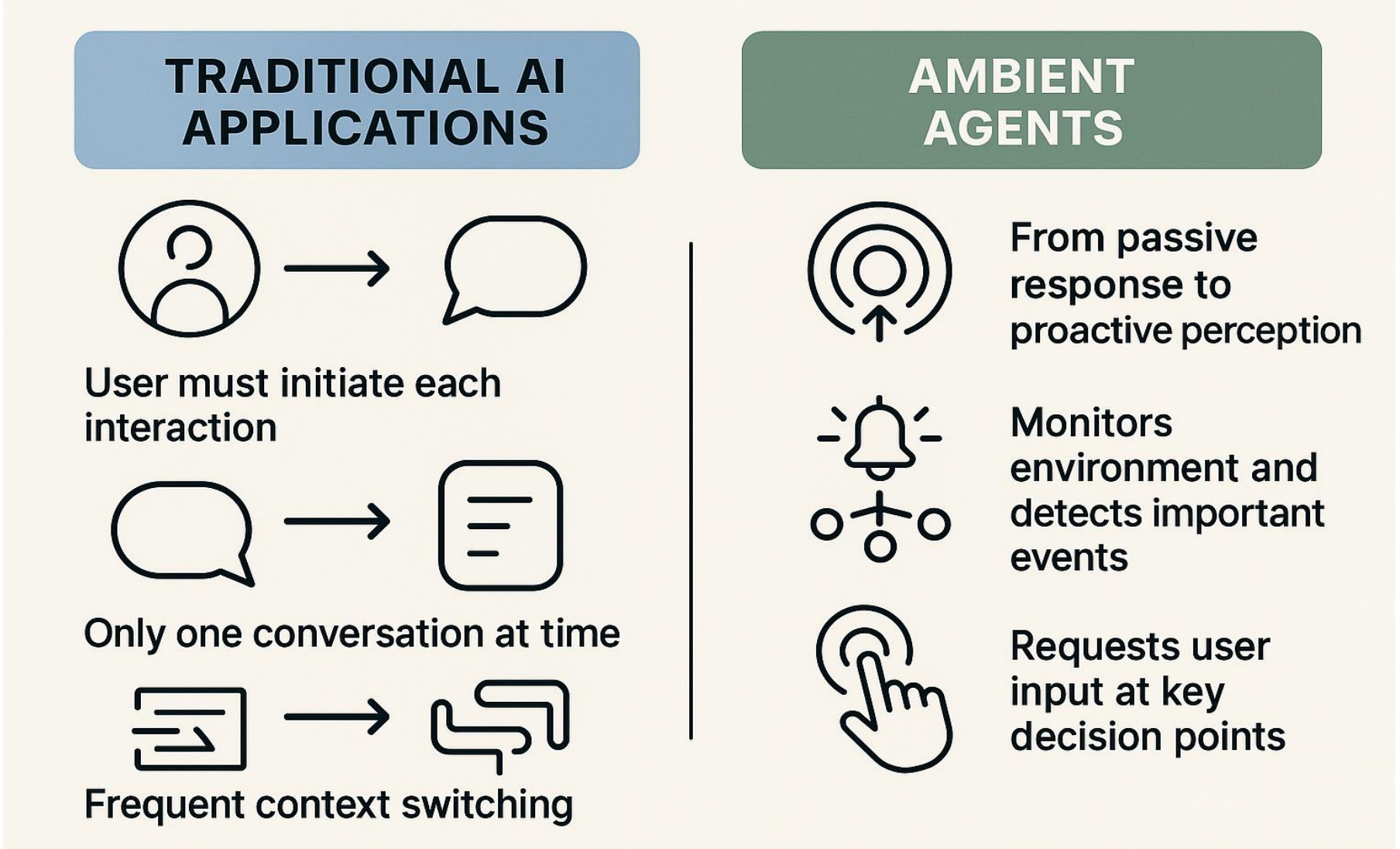
能够同时处理多个任务，而非传统聊天机器人的单一对话模式

减少交互负担

只在必要时才请求用户输入，最大限度减少对用户注意力的消耗

Ambient 智能体代表了 AI 应用从被动响应到主动感知的范式转变，通过事件驱动和多任务并行能力，显著减少用户交互负担。

图说 | 传统 AI 与 Ambient 智能体的区别



传统 AI 与 Ambient 智能体的区别

传统AI模式

- 用户必须主动发起每次交互
- 一次只能进行一个对话
- 用户需要频繁切换上下文
- 被动等待用户指令

Ambient智能体模式

- 从被动响应到主动感知
- 能够自主监控环境并识别重要事件
- 支持多个智能体同时运行
- 只在关键决策点请求用户参与

适用企业场景

电子邮件管理

日程安排

文档处理

社交媒体管理

研发流程

PART 02

LangGraph 智能体开放框架

LangGraph 框架介绍

- LangGraph 是一个用于构建基于图结构的代理应用。它允许开发者将复杂的 AI 任务分解为相互连接的节点和边，每个节点代表一个特定的操作或决策点。
- LangGraph 支持循环和分支逻辑，提供内置的状态持久化机制，并支持人机交互工作流。该框架的设计理念是使 AI 工作流更加模块化、可视化和可控。
- 它可独立使用，或与 LangChain 和 LangSmith 无缝集成，为开发者提供了强大的工具来创建高度灵活和可扩展的语言处理系统。

- **优势**：比较全面地支持循环和分支，提供内置持久性，支持人机交互工作流；与 LangChain 生态系统集成非常良好，也可以独立使用。
- **劣势**：学习曲线可能较陡；在简单任务上可能显得过于复杂。
- **机会**：在需要复杂决策树的应用中有巨大潜力，可以与其他 AI 工具和框架集成，增强功能。
- **威胁**：可能面临来自更简单、直观框架的竞争；图结构的复杂性可能影响性能，特别是在大规模应用中。

```python
from langgraph.graph import StateGraph, END
from langgraph.prebuilt import ToolNode

def should_continue(state):
    messages = state['messages']
    last_message = messages[-1]
    if last_message.tool_calls:
        return "tools"
    return END

def call_model(state):
    messages = state['messages']
    response = model.invoke(messages)
    return {"messages": [response]}

workflow = StateGraph(MessagesState)
workflow.add_node("agent", call_model)
workflow.add_node("tools", ToolNode(tools))
workflow.add_edge(START, "agent")
workflow.add_conditional_edges("agent", should_continue)
workflow.add_edge("tools", 'agent')

app = workflow.compile()
final_state = app.invoke(
  {"messages": [HumanMessage(content="what is the weather in sf")]}
)
```

图说｜业内一些知名智能体框架的关注点

AutoGen

专注于多代理对话应用

CrewAI

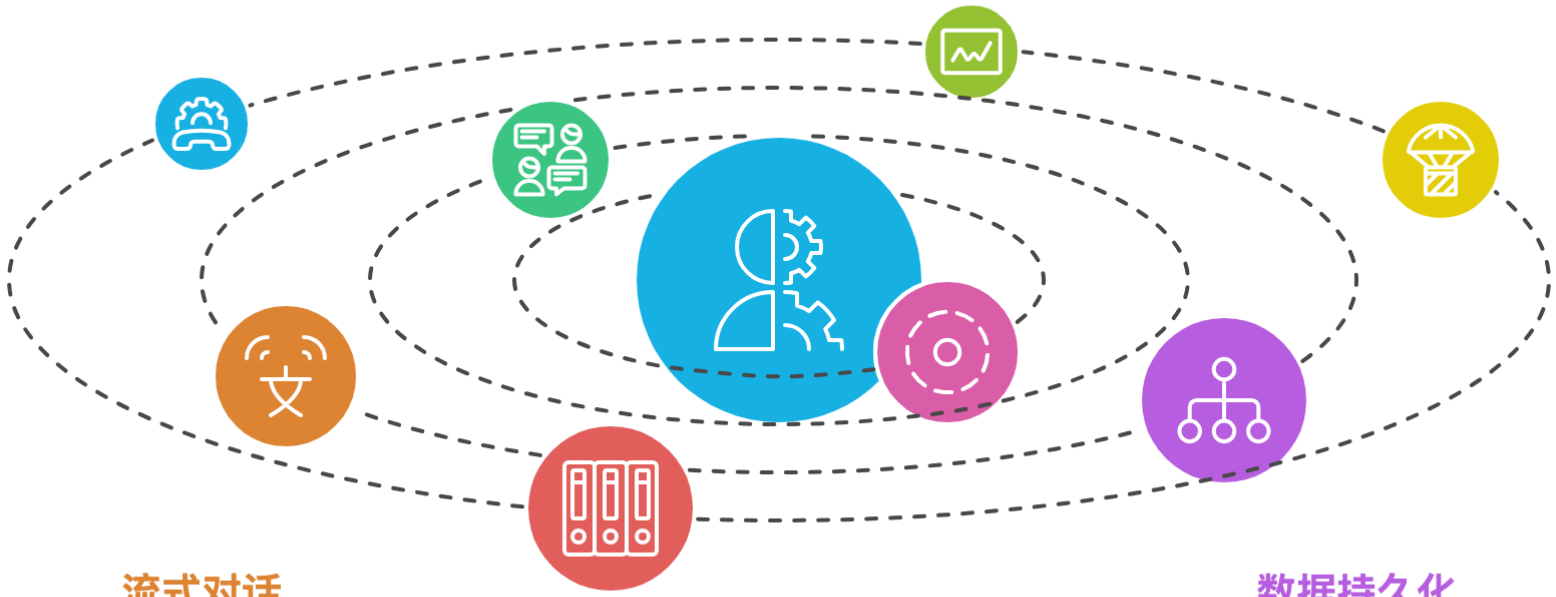
强调角色扮演和团队协作

LangGraph

基于图结构的复杂任务

Llama Deploy

设计用于异步和分布式部署



流式对话

提供多样化的输出选项

会话记忆

提供灵活的存储解决方案

人机交互

高级交互功能

数据持久化

在图结构持久化方面表现突出

LangGraph 强力支持 Ambient 智能体的构建

持久化层与状态管理

在每个操作或图节点之间保存代理状态，允许智能体在等待用户反馈时"暂停"执行，支持短期对话记忆和上下文保持，确保系统崩溃或重启后能恢复执行状态。

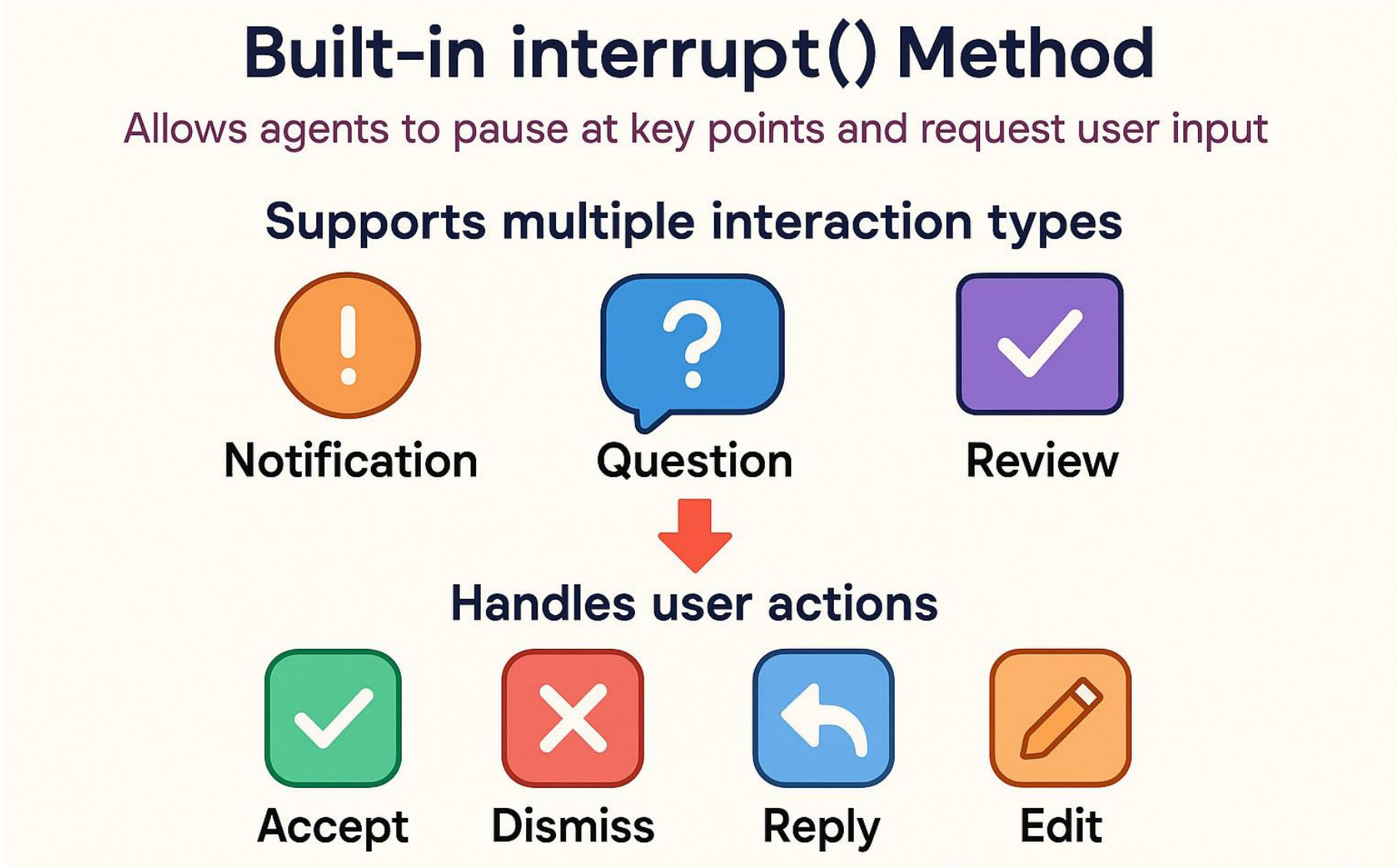
人机协作支持机制

提供内置的 interrupt 方法，允许智能体在关键点暂停并请求用户输入，支持多种交互类型（通知、提问、审查），能够处理用户的接受、忽略、回复或编辑操作。

长期记忆与定时任务

提供命名空间化的键值存储，支持语义搜索，允许智能体根据用户反馈更新记忆，促进智能体随时间学习和适应。内置 cron 作业支持，允许智能体按计划运行。

图说｜LangGraph 的人机环路



代码 | 为 Ambient 智能体设计的中断结构体

human-in-the-loop Python

```python
# 示例：使用 interrupt 方法请求用户输入
request: HumanInterrupt = {
    "action_request": {
        "action": "draft_email",
        "args": {"recipient": "client@example.com", "subject": "项目更新"}
    },
    "config": {
        "allow_ignore": True,
        "allow_respond": True,
        "allow_edit": True,
        "allow_accept": True
    },
    "description": "请审核这封电子邮件草稿"
}
response = interrupt([request])[0]
```

PART 03

企业级人机环路工作流程设计

企业级 Ambient 智能体的主要交互方式

通知模式

提醒用户注意重要事件，但不代表用户采取行动。适用于重要邮件到达、紧急会议请求、截止日期提醒等场景。

提问模式

获取缺失信息，解除智能体决策阻塞。适用于需要用户偏好的决策、缺少关键信息的任务、多个可能选项需要用户选择的情况。

审查模式

请求用户审核智能体生成的内容或决策。适用于邮件草稿审核、会议安排确认、文档审批等需要人类最终确认的场景。

企业级 Ambient 智能体的主要交互方式：通知模式

通知模式特点

提醒用户注意重要事件，但不代表用户采取行动。

适用场景：

- 重要邮件到达（如包含DocuSign的邮件）
- 紧急会议请求
- 截止日期提醒
- 异常事件检测

> *"通知：让用户知道某些事件很重要，但不采取任何行动。这在标记用户应该看到但代理无权处理的事件时很有用。"*

实现方式

```
{
    "action_request": {
        "action": "important_notification",
        "args": {"type": "docusign", "sender": "legal@company
    },
    "config": {
        "allow_ignore": true,
        "allow_respond": false,
        "allow_edit": false,
        "allow_accept": false
    },
    "description": "检测到来自法律部门的DocuSign请求，需要您的签名。
}
```

用户体验考量：

- ✓ 通知应简洁明了
- ✓ 提供足够上下文让用户理解重要性
- ✓ 允许用户轻松忽略或标记为已读

企业级 Ambient 智能体的主要交互方式：提问模式

提问模式特点

获取缺失信息，解除智能体决策阻塞。

适用场景：

- 需要用户偏好的决策（如是否参加会议）
- 缺少关键信息的任务
- 多个可能选项需要用户选择
- 需要确认假设的情况

> *"提问：向用户提问以帮助解除代理的阻塞。代理可能正在尝试采取一些行动，但不清楚如何最好地执行，因为它缺乏一些相关信息。"*

实现方式

```
{
    "action_request": {
        "action": "conference_attendance",
        "args": {
            "conference": "AI Summit 2023",
            "dates": "Oct 15-17",
            "location": "San Francisco"
        }
    },
    "config": {
        "allow_ignore": true,
        "allow_respond": true,
        "allow_edit": false,
        "allow_accept": true
    },
    "description": "您收到了参加AI Summit 2023的邀请。您想要参加吗
}
```

用户体验考量：

- ✓ 问题应清晰具体
- ✓ 提供足够上下文帮助用户做出决策
- ✓ 支持简单的是/否回答和详细解释

企业级 Ambient 智能体的主要交互方式：审查模式

审查模式特点

让用户审核智能体计划执行的重要操作。

适用场景：

- 外发电子邮件
- 日历邀请创建
- 社交媒体发布
- 文档提交或签署
- 资源分配决策

> *"审查：审查代理想要采取的行动。有些行动足够'危险'，值得为代理想要采取的任何行动硬编码审查。"*

实现方式

```
{
    "action_request": {
        "action": "send_email",
        "args": {
            "to": "client@example.com",
            "subject": "项目提案",
            "body": "尊敬的客户，附件是我们的项目提案..."
        }
    },
    "config": {
        "allow_ignore": false,
        "allow_respond": true,
        "allow_edit": true,
        "allow_accept": true
    },
    "description": "请审核这封发给客户的项目提案邮件。"
}
```

用户体验考量：

- ✓ 提供完整的操作上下文
- ✓ 支持直接编辑内容
- ✓ 允许用户提供修改建议
- ✓ 对于高风险操作，可禁用忽略选项

Agent Inbox：集中管理智能体交互的新型界面

Agent Inbox 核心理念

Agent Inbox 是专为 Ambient 智能体设计的用户界面，用于跟踪和管理所有未完成的操作。

> *"我们转向了我们称为 Agent Inbox 的东西。这是与 Ambient 智能体交互的新 UX。它模仿了电子邮件收件箱和客户支持票务系统的某种组合。"*

这种设计解决了传统聊天界面的局限性，特别是在处理多个并行任务时的混乱问题。未来版本将支持基于优先级的排序和团队协作功能。

设计理念

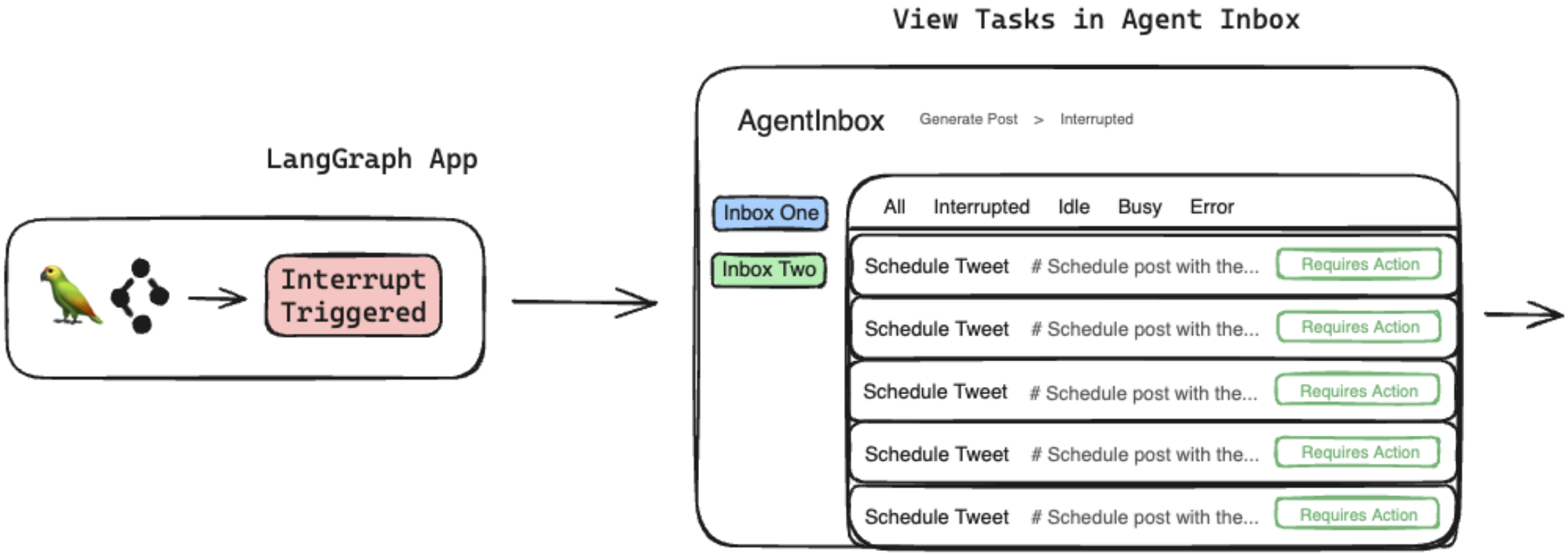
集中管理交互
将所有智能体交互集中在一个界面中

减少上下文切换
避免在多个聊天窗口间切换的认知负担

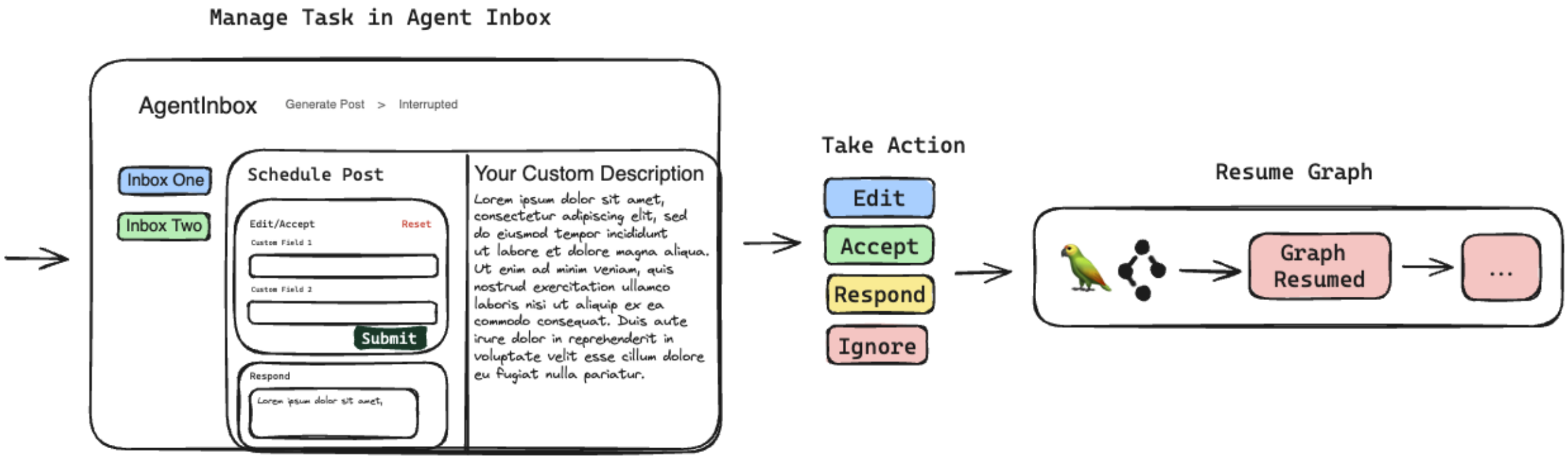
任务跟踪
清晰显示所有未完成的操作和请求

灵活的UI组件
支持添加专用面板、按钮和其他交互元素

图说 | Agent Inbox 界面操作流程 1/2



图说 | Agent Inbox 界面操作流程 2/2



PART 04

企业级智能平台的实践与思考

案例｜企业邮件助手智能体

邮件智能分类

- **忽略**：自动标记为已读
- **通知**：提醒用户但不自动回复
- **回复**：尝试起草回复内容

日历管理

- 查看用户日程
- 协助安排会议
- 创建日历事件（需用户批准）

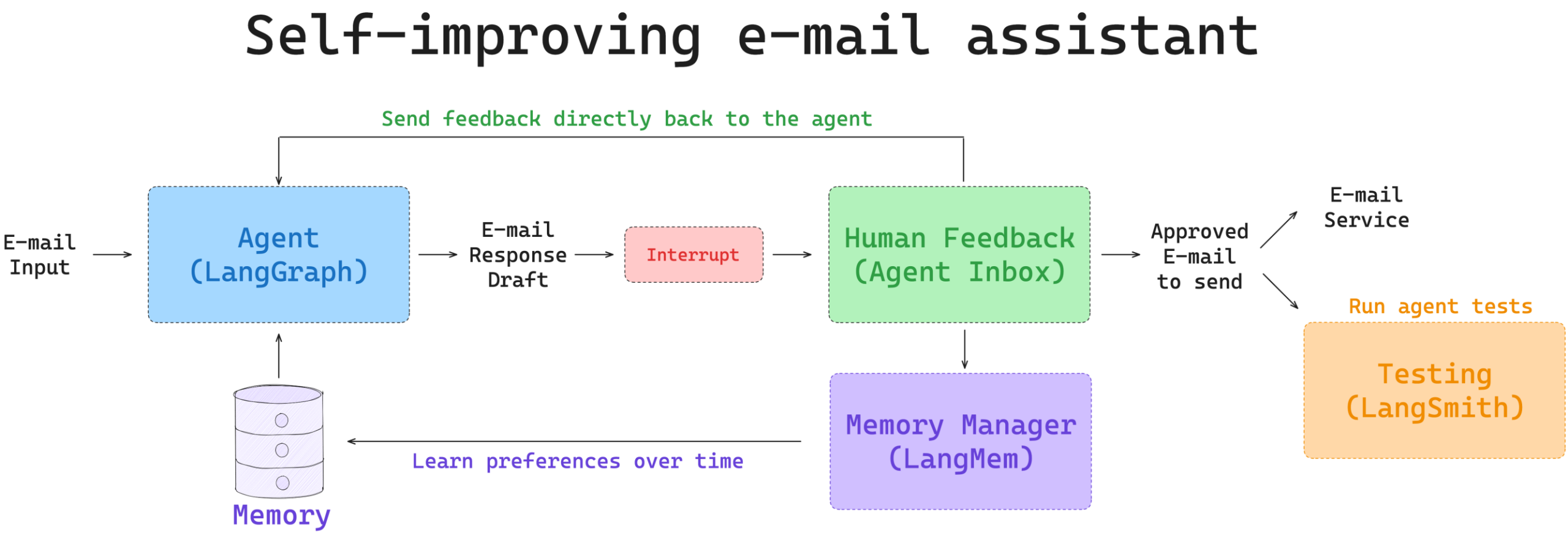
人机协作状态

- **中断状态**：需要用户介入
- **空闲状态**：已完成处理
- **忙碌状态**：正在处理
- **错误状态**：处理过程中遇到错误

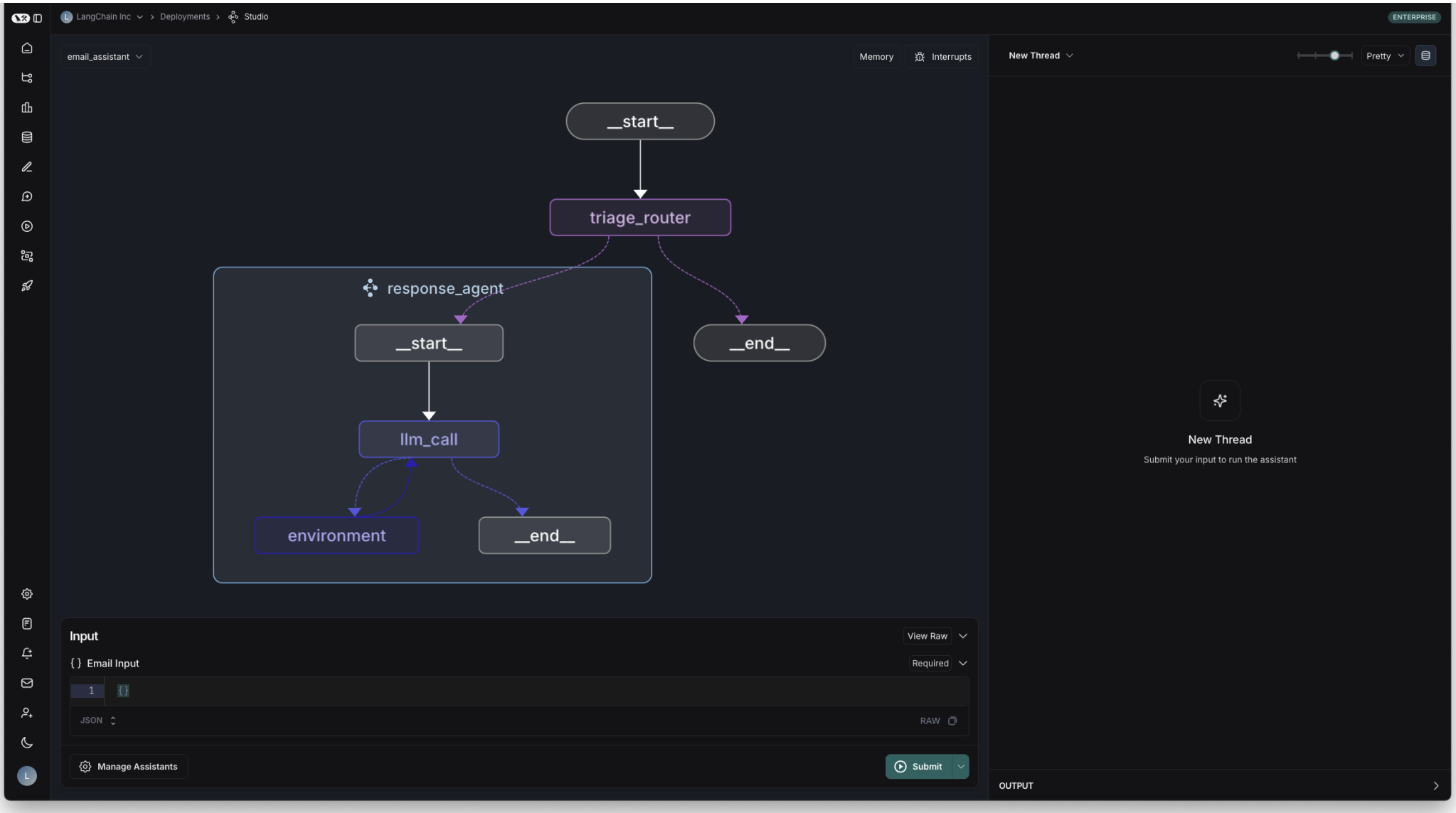
中断类型

- **通知型中断**：仅通知用户
- **问询型中断**：向用户提问
- **回复草稿中断**：请求审核邮件草稿
- **日程安排中断**：确认会议安排

案例｜企业邮件助手智能体（工作流程总览）



案例｜企业邮件助手智能体（智能体核心工作流）



案例｜企业邮件助手智能体（智能体系统提示词）

< Role >
You are a top-notch executive assistant who cares about helping your executive perform as well as possible.
</ Role >

< Tools >
You have access to the following tools to help manage communications and schedule:
1. triage_email(ignore, notify, respond) - Triage emails into one of three categories
2. write_email(to, subject, content) - Send emails to specified recipients
3. schedule_meeting(attendees, subject, duration_minutes, preferred_day) - Schedule calendar meetings
4. check_calendar_availability(day) - Check available time slots for a given day
</ Tools >

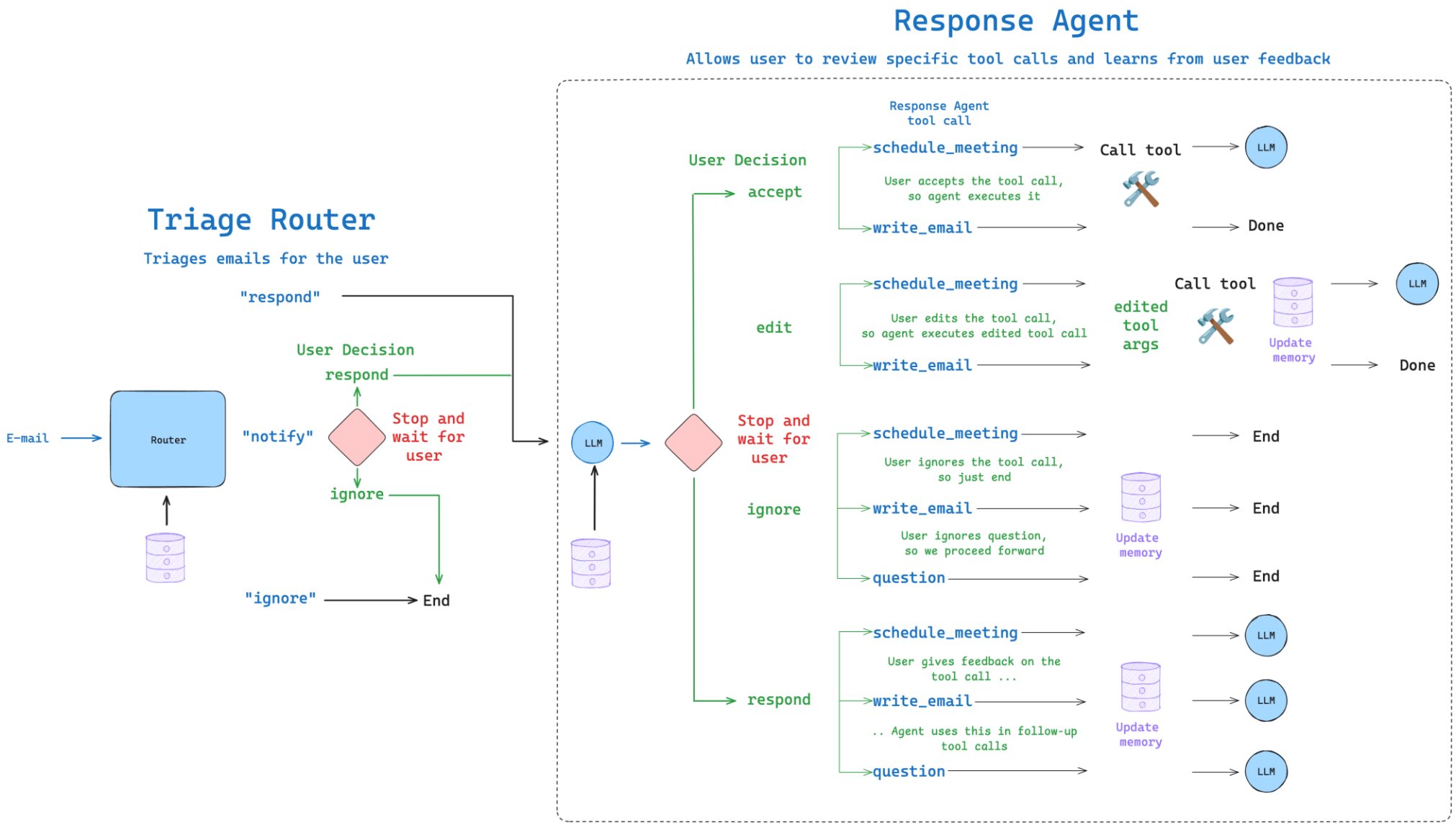
< Instructions >
When handling emails, follow these steps:
1. Carefully analyze the email content and purpose
2. For responding to the email, draft a response email with the write_email tool
3. For meeting requests, use the check_calendar_availability tool to find open time slots
4. To schedule a meeting, use the the schedule_meeting tool
5. If you scheduled a meeting, then draft a short response email using the write_email tool
6. After using the write_email tool, the task is complete
</ Instructions >

< Triage Instructions >
Emails that are not worth responding to:
- Marketing newsletters and promotional emails and spam

There are also other things that should be known about, but don't require an email response. For these, you should notify (using the `notify` response).
Examples of this include:
- Team member out sick or on vacation

Emails that are worth responding to:
- System Admin notifications (use a brief thank you for the email)
- Direct questions from team members requiring expertise
- Meeting requests requiring confirmation
</ Triage Instructions >

案例｜企业邮件助手智能体（智能体人机环路）

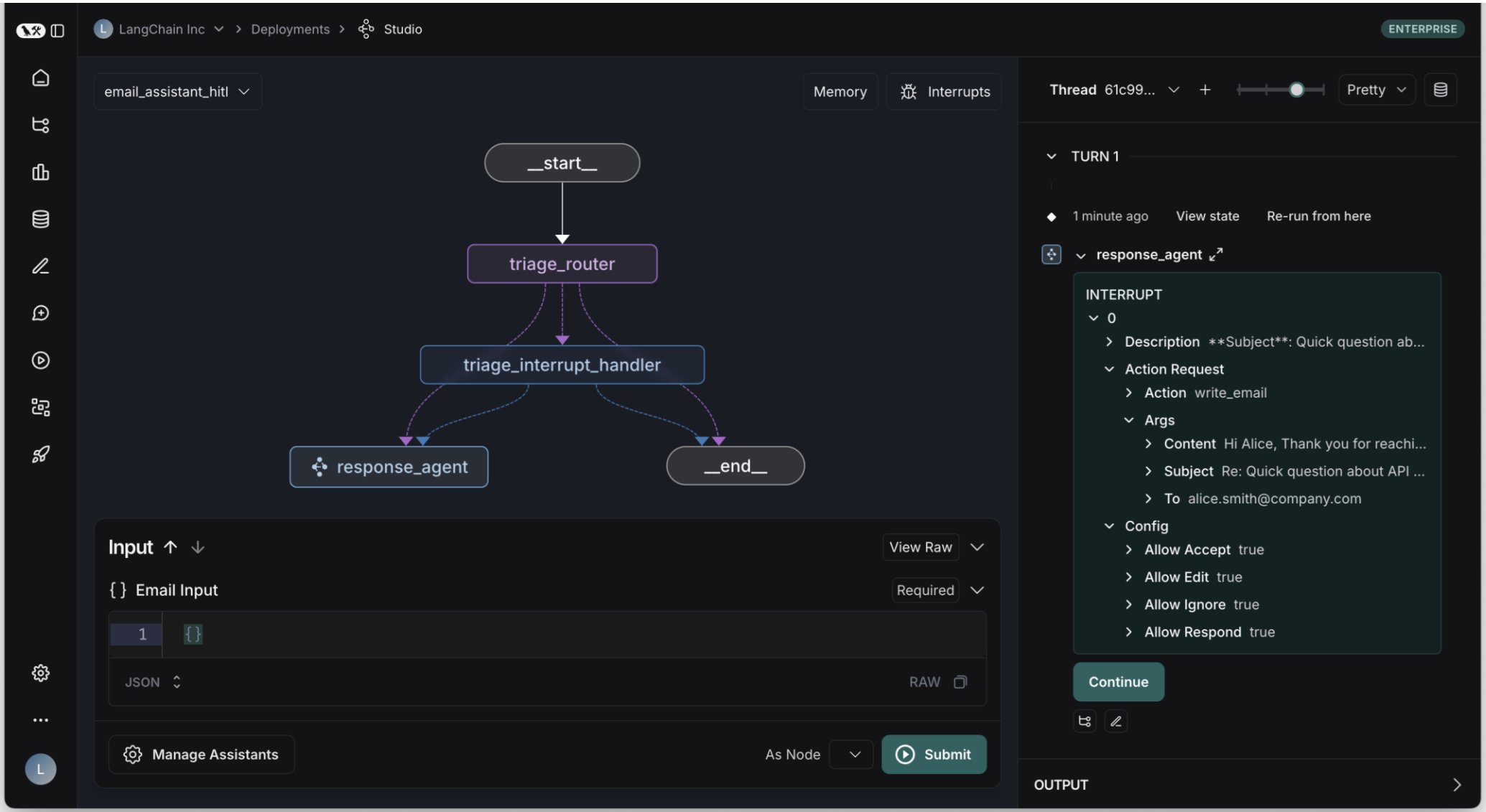


案例｜企业邮件助手智能体（人机环路结构体）

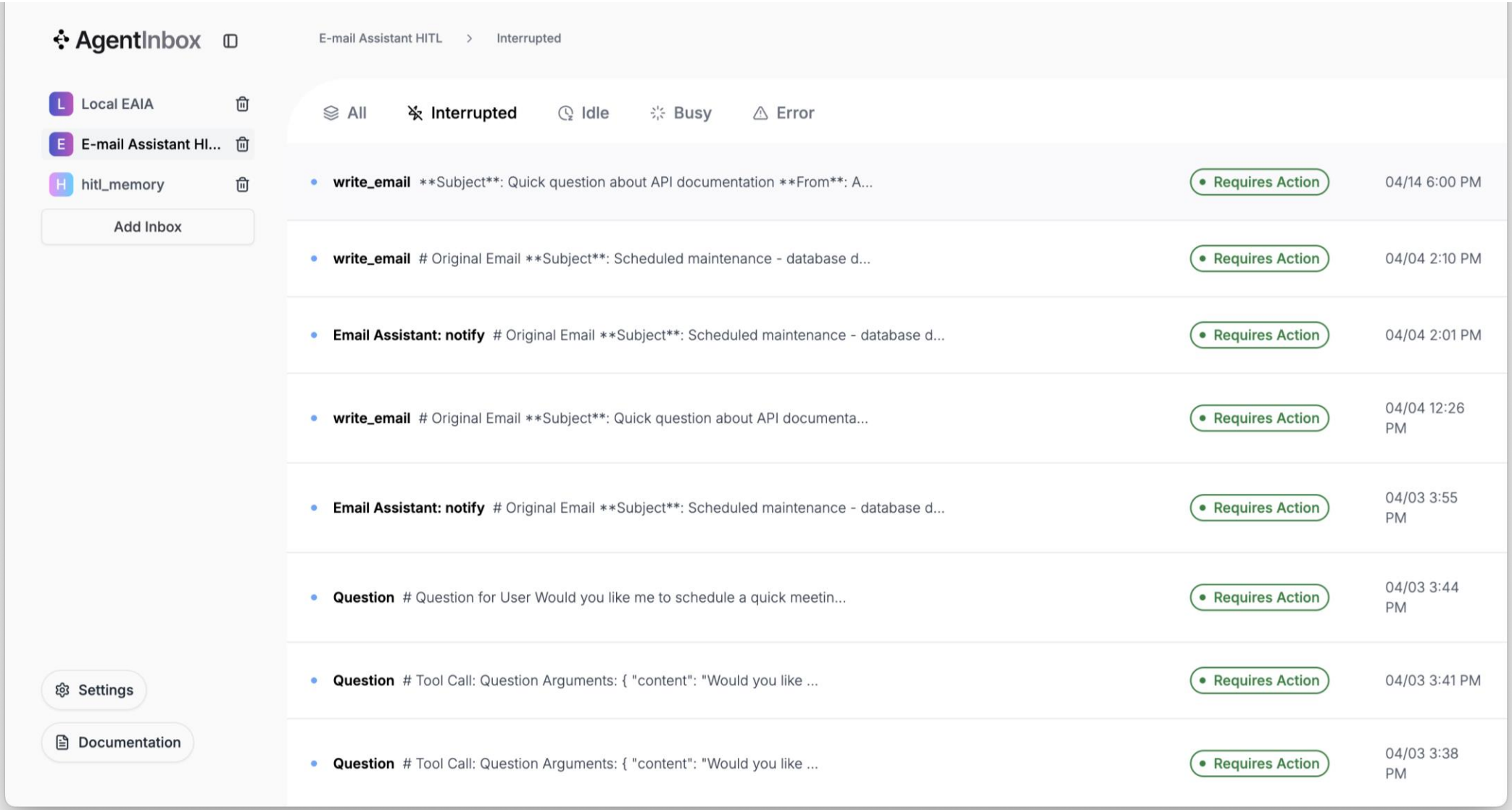
Email Assistant (Human-in-the-Loop) Python

```python
request = {
    "action_request": {
        "action": "write_email",  # Name of the tool to call
        "args": {"to": "john@example.com", "subject": "Meeting", "content": "..."}  # Action parameters
    },
    "config": {
        "allow_ignore": True,   # Can dismiss the action
        "allow_respond": True,  # Can provide feedback
        "allow_edit": True,     # Can modify the action
        "allow_accept": True,   # Can approve the action
    },
    "description": "Email content to display..." # Context shown to the user
}
```

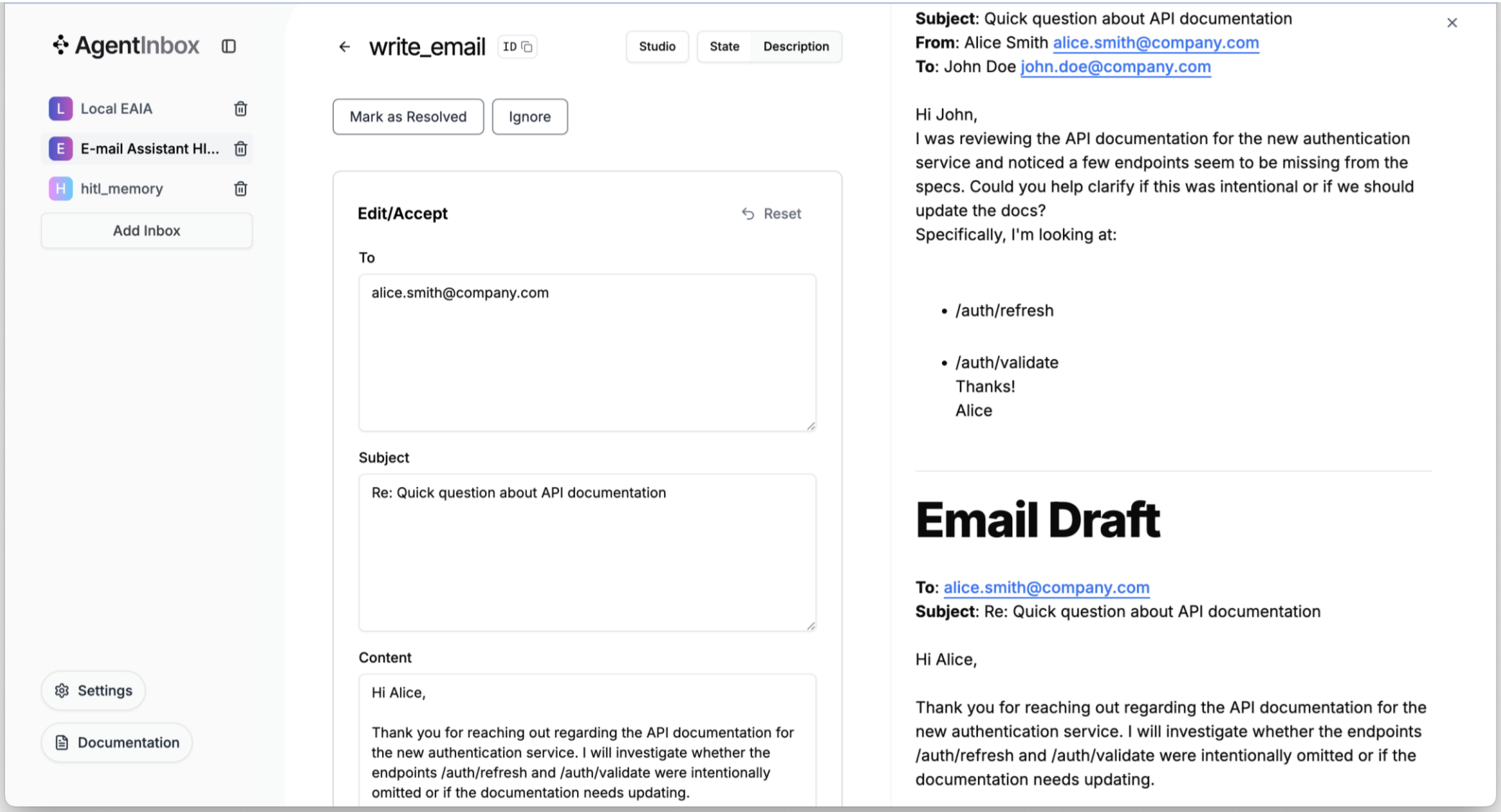
案例｜企业邮件助手智能体（含人机环路后的工作流）



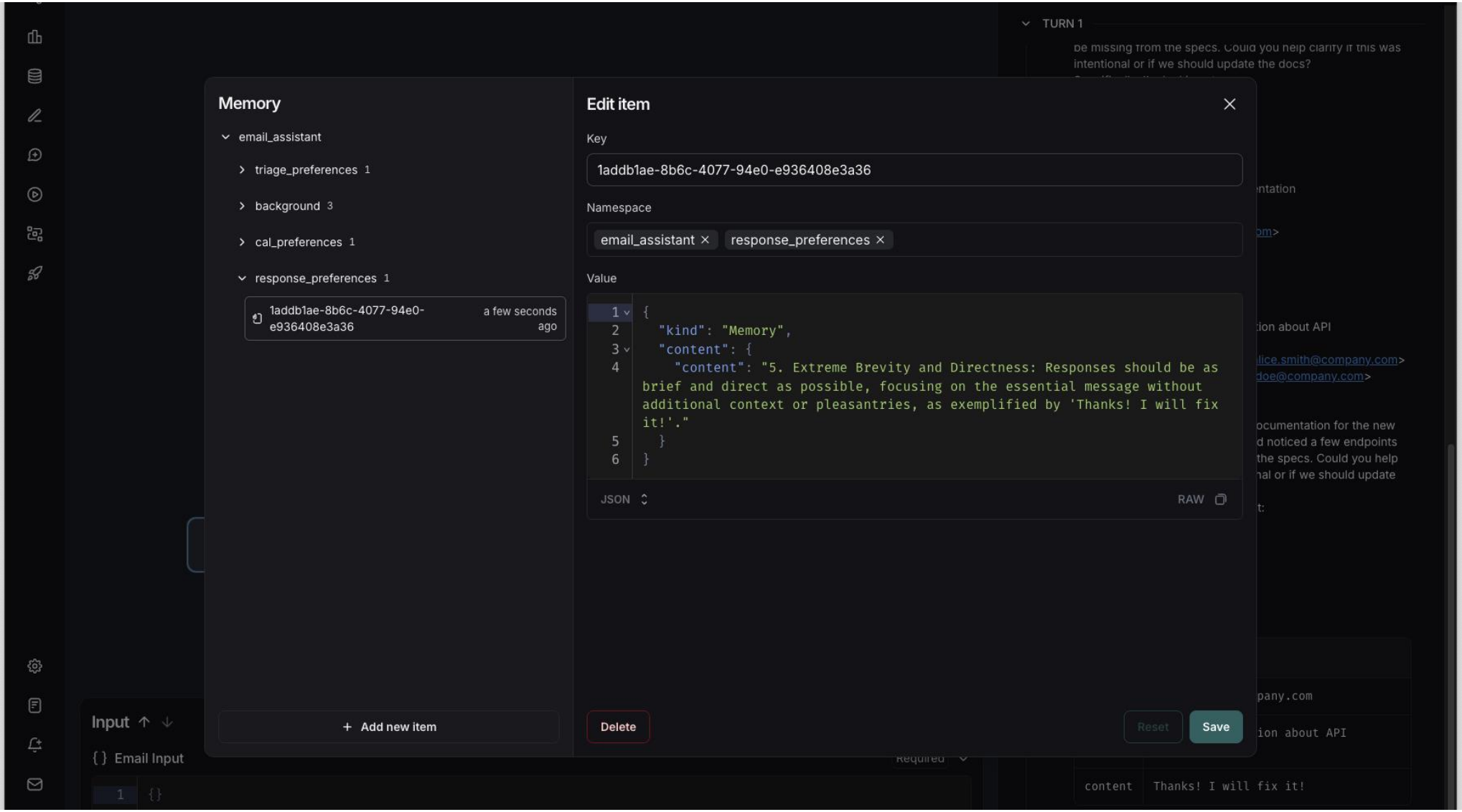
案例｜企业邮件助手智能体（Agent Inbox 待办箱）



案例｜企业邮件助手智能体（Agent Inbox 用户处理）



案例｜企业邮件助手智能体（用户处理后更新记忆）



案例衍生｜企业审批流程助手智能体

基于 LangGraph 的审批请求管理工作流

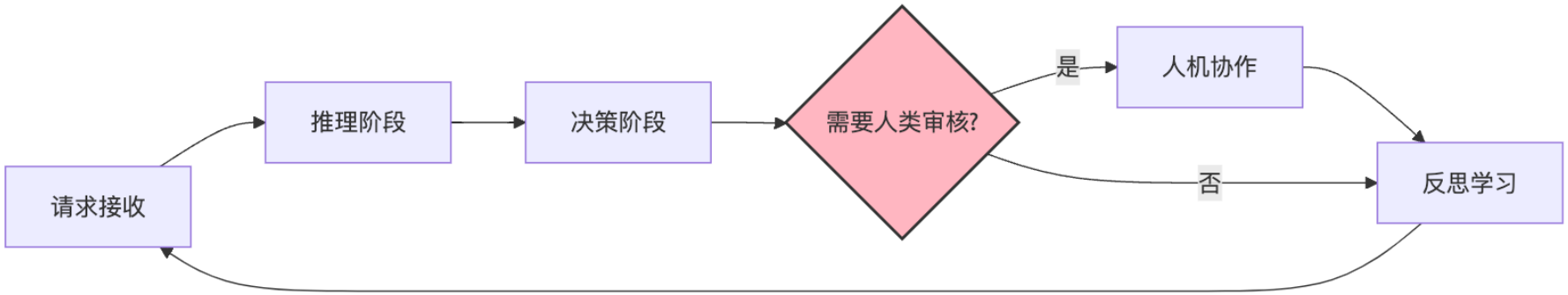
构建完整的审批流程图，实现请求接收、推理、决策和反思等状态间的精确转换

动态提示组合处理多样化审批请求

根据不同类型的审批请求，动态生成适合的提示模板，提高处理精度

通过反思机制持续学习和改进

记录人类审核修改的请求、解释和最终决策，对错误决策进行结构化反思



🤔 思考｜决定何时引入人类参与

风险评估框架

- **高风险操作**：涉及财务、法律或重要客户关系的决策
- **中等风险**：可能影响效率但不会造成严重后果的操作
- **低风险**：常规、重复性任务，出错后果有限

不确定性阈值

- 当智能体置信度低于预设阈值时请求人类参与
- 使用统计方法量化决策不确定性
- 动态调整阈值基于历史表现

新颖性检测

- 识别与历史数据显著偏离的情况
- 对于首次遇到的场景自动请求人类指导
- 建立新颖场景库以减少未来干预

思考 | 设计有效的人机协作界面

上下文丰富性

- 提供足够背景信息使决策高效
- 包含相关历史数据和先前决策
- 突出关键信息点减少认知负担

交互效率

- 最小化完成交互所需的点击/操作
- 支持快捷键和批量操作
- 提供默认选项和智能建议

透明度与可解释性

- 清晰解释为何需要人类参与
- 展示智能体的推理过程
- 提供决策依据和可能的替代方案

学习机制

- 记录人类反馈以改进未来决策
- 实现渐进式自动化（随时间减少干预）
- 提供反馈循环让用户了解系统如何学习

🤔 思考 | 实现渐进式自动化的策略

学习曲线设计

- 初始阶段：高度人类参与，系统主要学习
- 中间阶段：选择性人类参与，系统逐步接管常见任务
- 成熟阶段：最小人类参与，仅处理例外情况

用户信任建立

- 从低风险任务开始自动化
- 提供透明的性能指标和改进证据
- 允许用户调整自动化程度

反馈循环优化

- 收集结构化反馈改进模型
- 定期审核自动化决策质量
- 建立明确的升级路径处理失败案例

🤔 思考 | 优化 Ambient 智能体性能的关键策略

模型优化

- 选择适当的模型大小平衡性能和资源
- 实现模型量化减少内存占用
- 使用模型蒸馏技术提高效率
- 优化推理批处理提高吞吐量

缓存策略

- 实现多层缓存架构
- 缓存常见查询和响应
- 使用预计算减少实时计算
- 实现智能缓存失效机制

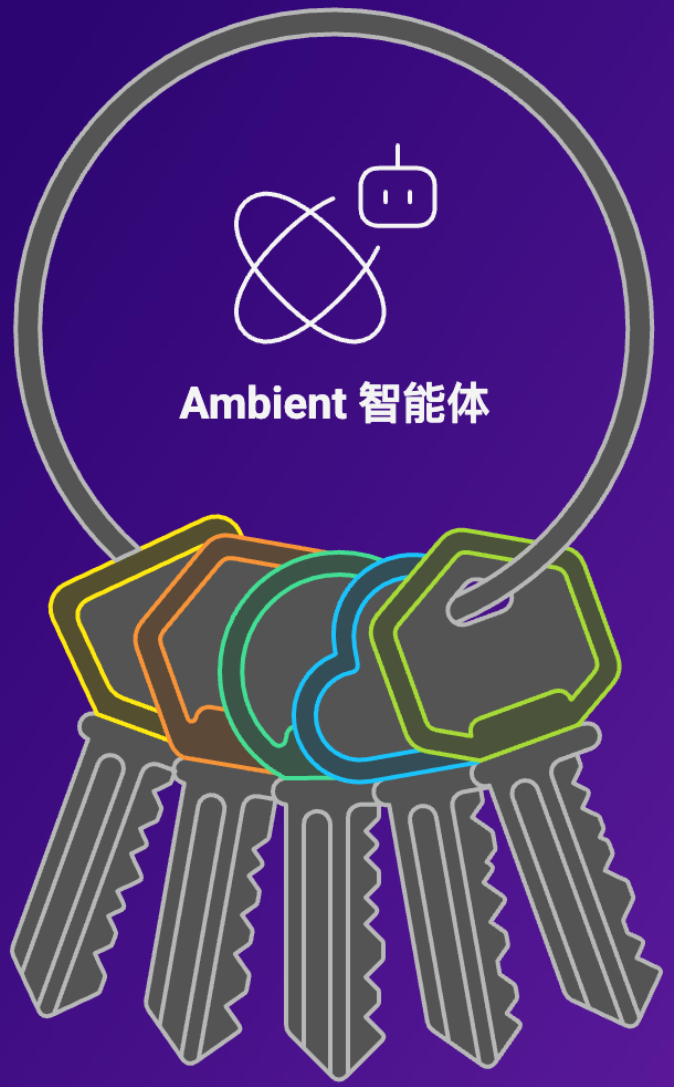
资源管理

- 实现资源池化和重用
- 优化内存管理减少垃圾收集
- 实施资源限制防止过度使用
- 监控和分析资源使用模式

监控与优化循环

- 实施全面性能监控
- 建立性能基准和目标
- 定期性能审查和瓶颈分析
- 持续优化反馈循环

创新 Ambient 智能体框架与人机协作



事件驱动

智能体响应环境中的事件和信号，无需用户主动发起。

多任务并行

智能体能够同时处理多个任务，提高效率。

减少交互负担

智能体最大限度地减少用户交互，只在必要时请求输入。

持久化层

智能体在操作之间保存状态，确保连贯性和恢复能力。

人机协作

智能体与用户协作，在关键决策点请求输入。

AiDD 6th 2025

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！

扫码参与调研

2025 AI+研发数字峰会

拥抱AI 重塑研发

科技生态圈峰会 + 深度研习

——1000 + 技术团队的共同选择

K+峰会

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

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

AiDD 6th 2025

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