

tRPC-Agent

腾讯多智能体框架的架构设计与业务实践

演讲人：陈明杰

腾讯 / 专家开发工程师

AiCon
全球人工智能开发与应用大会

目录

- 01 tRPC-Agent 开发背景
- 02 tRPC-Agent选型
- 03 tRPC-Agent架构
- 04 tRPC-Agent技术挑战
- 05 业务实践
- 06 未来展望

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

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

1000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

上海

1200人

01 tRPC-Agent开发背景

什么是tRPC

腾讯应用最广微服务框架

服务数6.4w+, 节点数260w+, 覆盖所有BG
超2000人参与共建

插件化

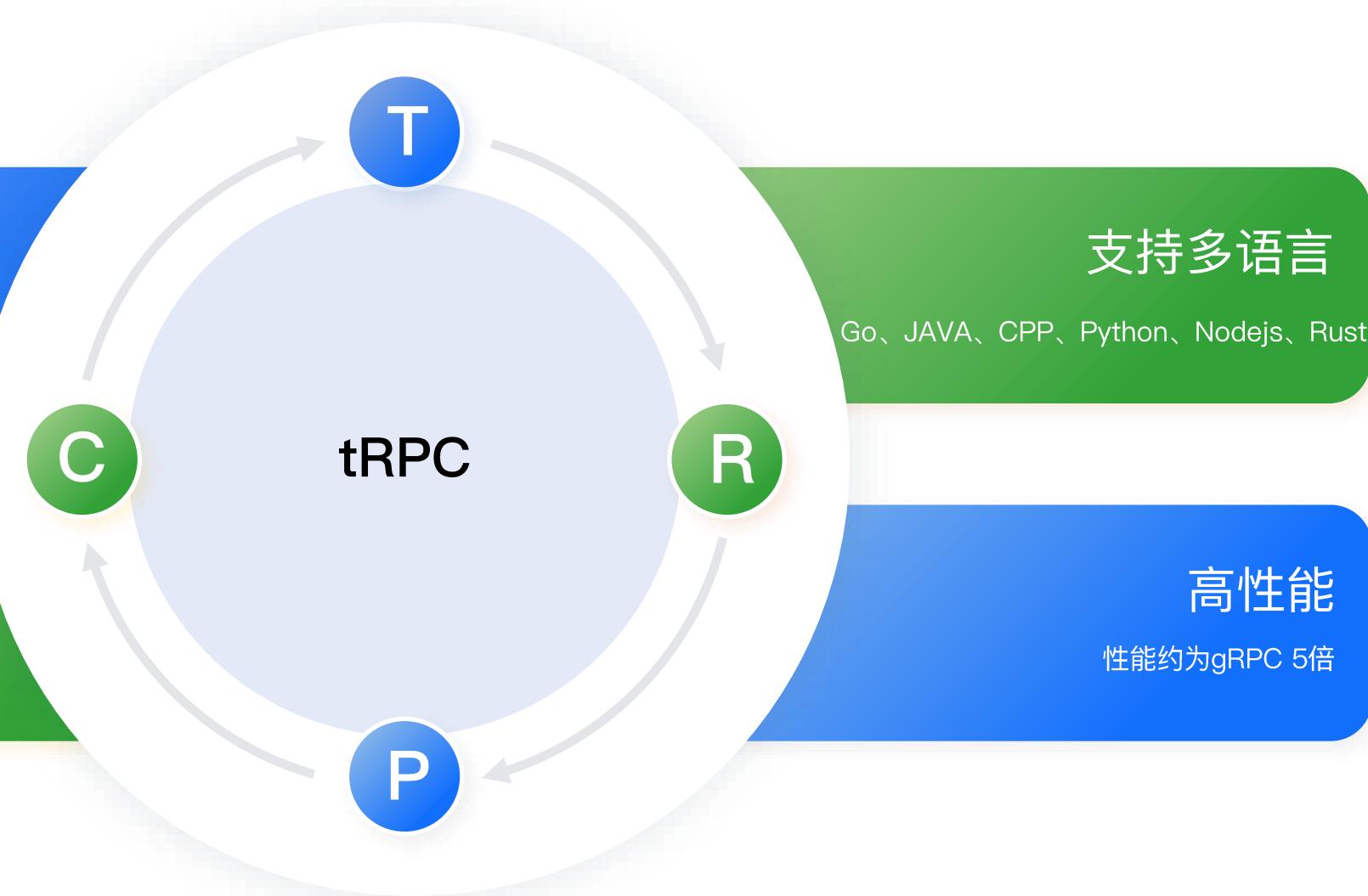
包含数据库、可观测、协议、名字服
务、配置等200+插件生态

支持多语言

Go、JAVA、CPP、Python、Nodejs、Rust

高性能

性能约为gRPC 5倍



为什么开发tRPC-Agent



02 tRPC-Agent选型



tRPC-Agent技术选型

自主多Agent协作框架

- 体现Autonomous Agent理念
- 环境感知、自主决策、动作执行
- 多Agent分布式协作
- 智能涌现特性

Workflow编排框架

- 工作流或者图、状态机和LLM交互
- 执行路径较为确定的场景

框架名称	类型	核心范式/理念	主要特点与定位
AutoGen	自主多Agent协作框架	对话驱动	将复杂任务建模为多个专业Agent间的自由对话，通过互动实现涌现式协作，高度灵活自主。
ADK	自主多Agent协作框架	Agent生命周期管理与编排	提供完整的Agent开发工具链，专注于多Agent的生命周期管理和协作编排。
CrewAI	自主多Agent协作框架	基于角色	模拟人类团队，通过为Agent分配明确的角色、目标和任务，实现面向流程的确定性协作。
Agno	自主多Agent协作框架	轻量高性能 & 多模态	轻量级高性能框架，专注于增强Agent的多模态能力和团队协作编排。
LangChain	workflow编排框架	链式抽象	通过预定义的链来组织和编排LLM调用及工具使用，构建应用的基础流行框架。
LangGraph	workflow编排框架	状态机/图结构	基于有状态图构建工作流，提供对复杂流程的极致控制、可靠性和可观测性。

结论：

腾讯业务场景复杂，对多Agent协作有强诉求，同时部分场景需要稳定的工作流输出保结果确定性。
tRPC-Agent的核心架构基于自主多Agent协作，同时提供工作流编排的能力。

03 tRPC-Agent技术架构

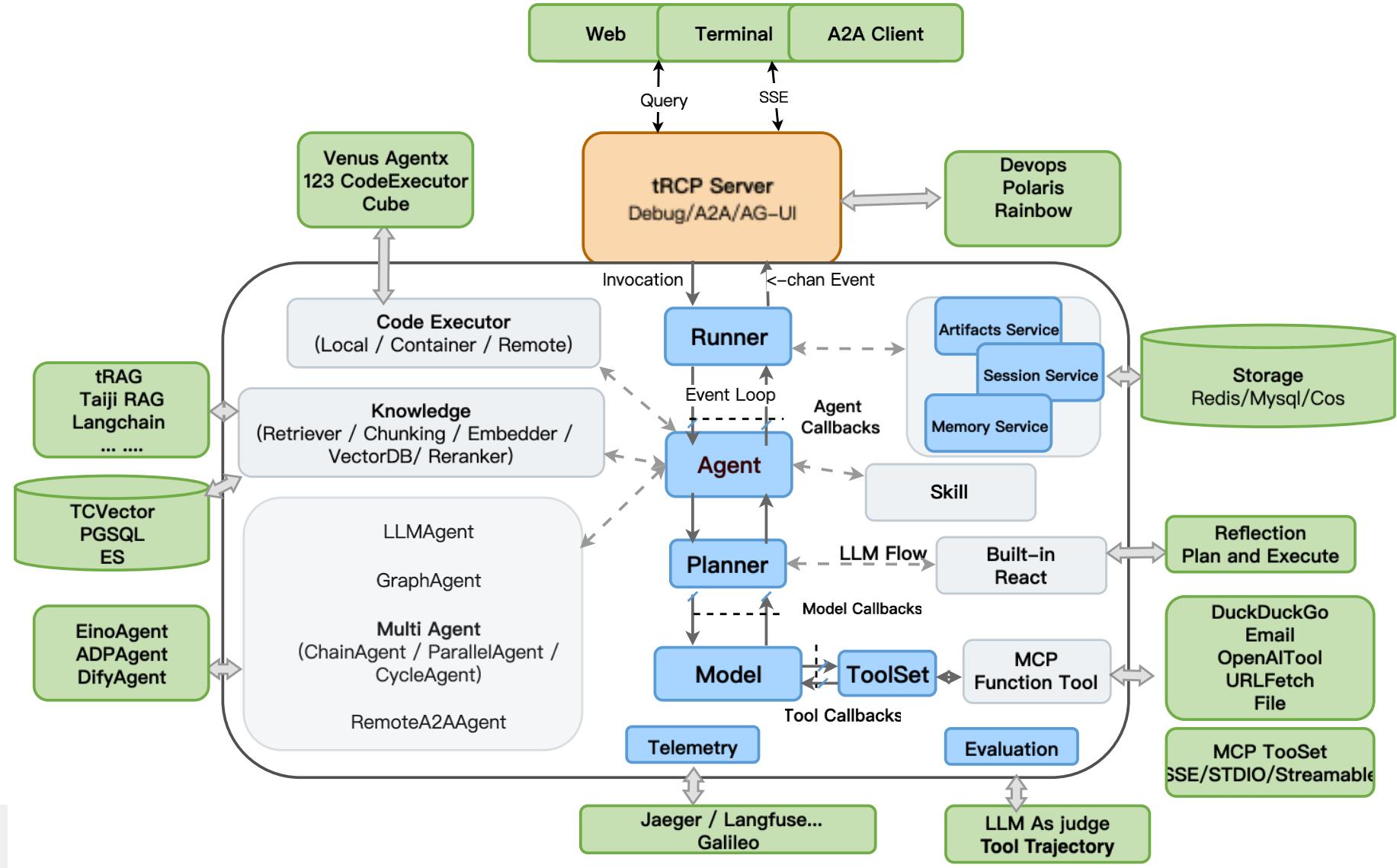
tRPC-Agent架构

挑战:

- Agent框架内部模块如何交互设计，方便扩展，并让用户使用简单；
- Agent框架如何复用tRPC已有生态运营体系，对接公司已有生态；

解决方案:

- 分层并模块化，采用插件工厂设计模式，支持Callback
各模块统一使用事件机制Event进行交互，进行解耦
Runner负责管理Agent生命周期
- tRPC与Runner结合快速构建tRPC服务，复用tRPC生态运营



04 tRPC-Agent技术挑战

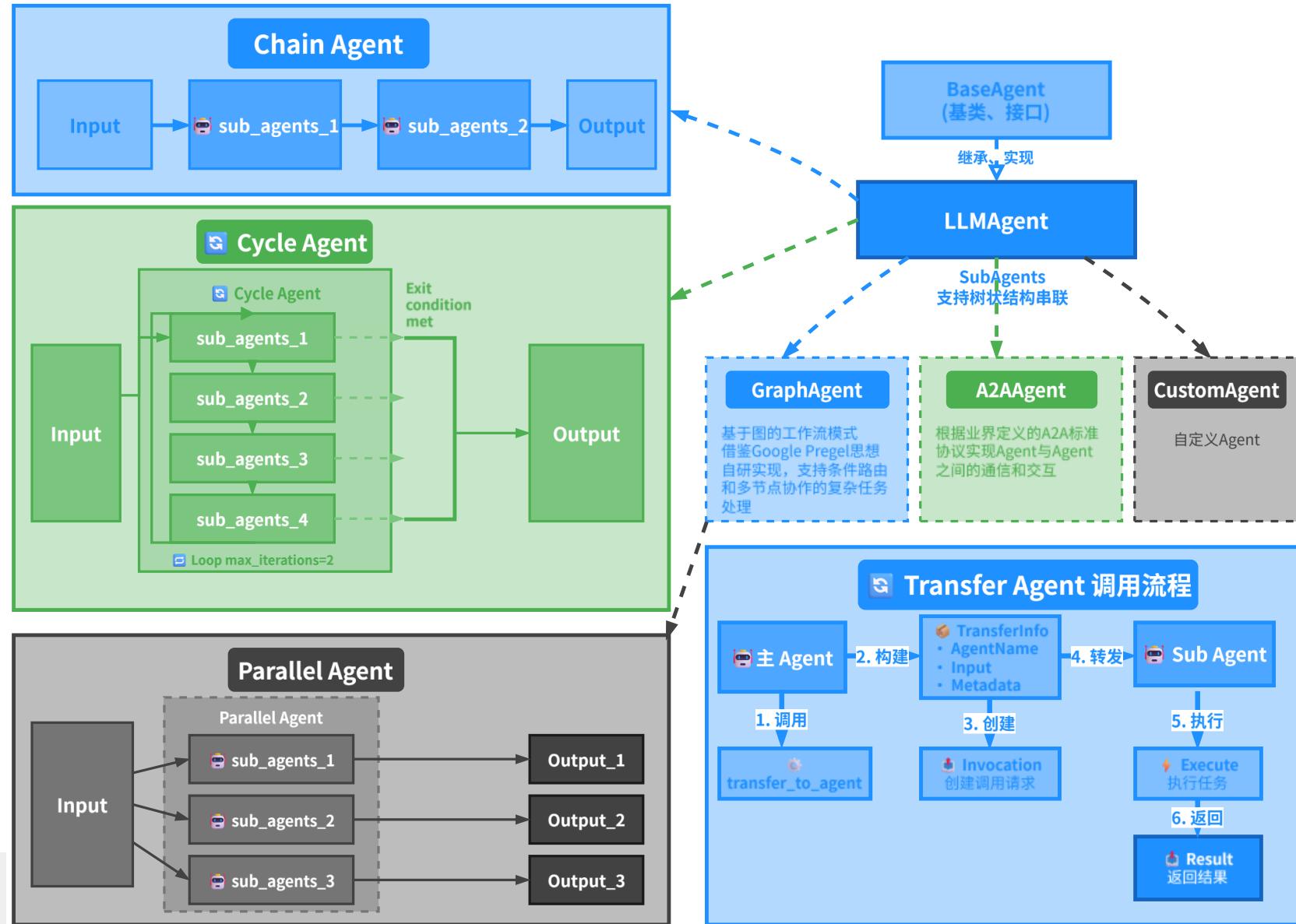
Multi-Agent-System

挑战：

- ✓ 如何支持多Agent编排和协作
- ✓ 满足各式各样的Agent需求

方案：

- ✓ 基于接口，实现多种Agent，内置常用多Agent组合链式、循环、并行多Agent协作；
- 基于Subagent机制，采用Transfer_to_agent工具调用实现Agent控制权转移，支持树状编排
- ✓ 可自定义Agent



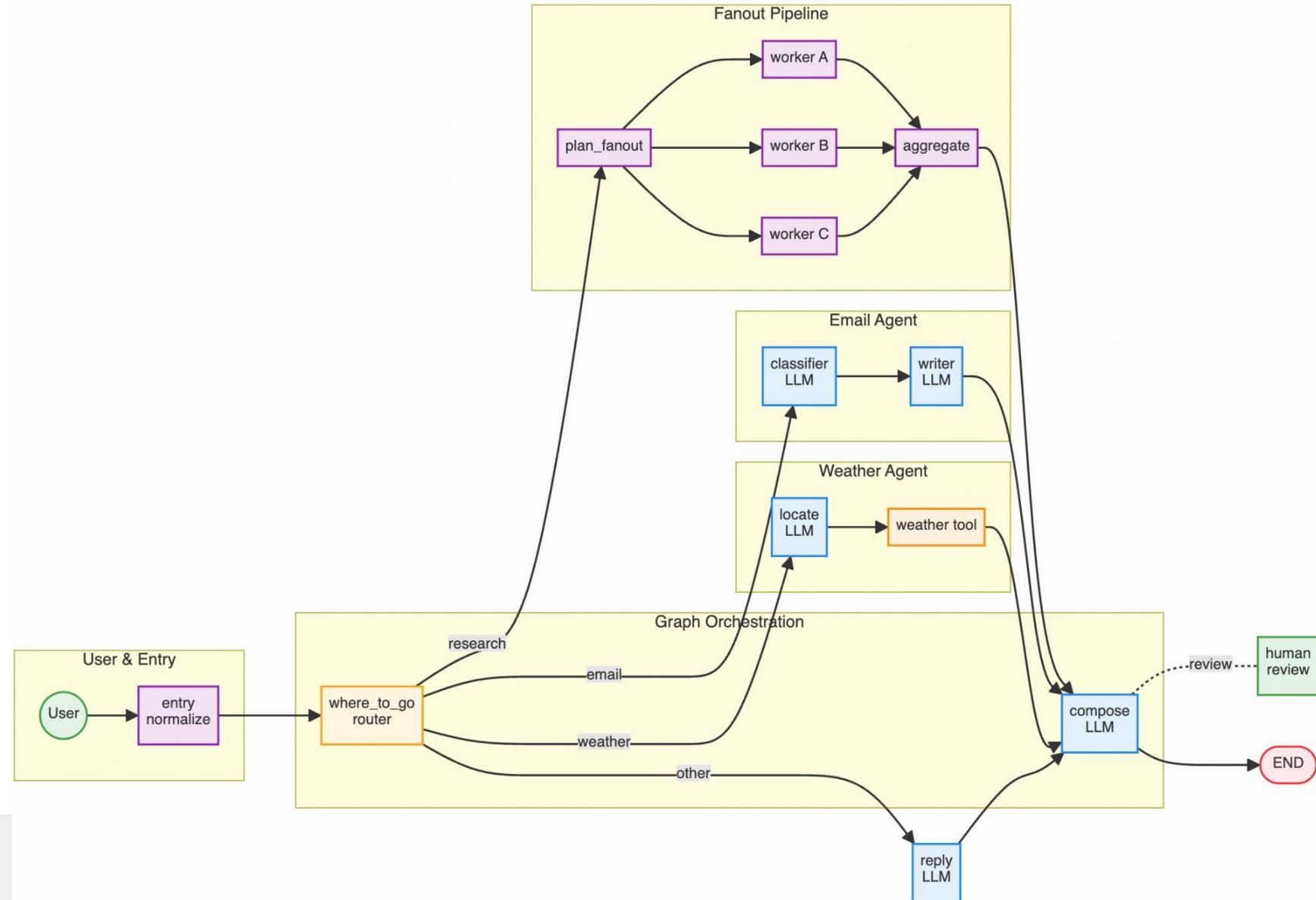
GraphAgent

挑战:

- 复杂的并行/串行混合执行
- 长任务稳定性、Human-in-loop
- 和现有Muti-Agent-System 结合

方案:

- Pregel BSP 执行模型
- 实现Checkpoint, 支持任务中断恢复、Human-in-Loop、时间旅行
- 基于Graph执行引擎封装
GraphAgent, 并通过Agent node支持
SubAgent串联



Planner

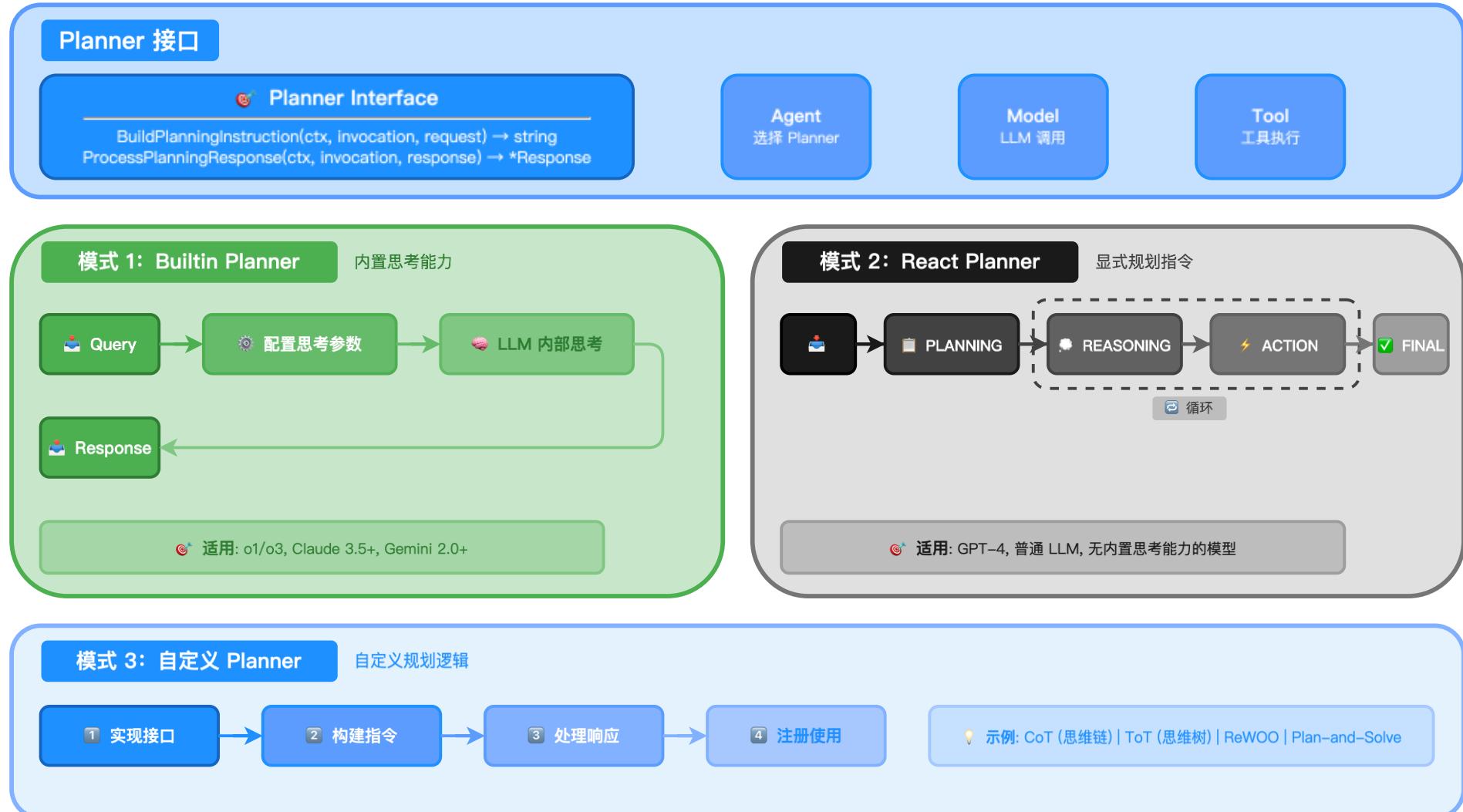
挑战：

- 如何让不同模型的Agent都具备自主性？

方案：

- 内置思考的built-in planner，开启模型思考
- 支持React planner，React 通过显式指令引导模型按标签格式输出，流式累积完整消息后再正则匹配标签

支持自定义Planner



MCP

挑战：

- 早期Go语言没有官方MCP SDK
- SSE版本协议缺陷，多节点会存在会话丢失
- tRPC 微服务用户需要快速开发MCP服务

解决方案：

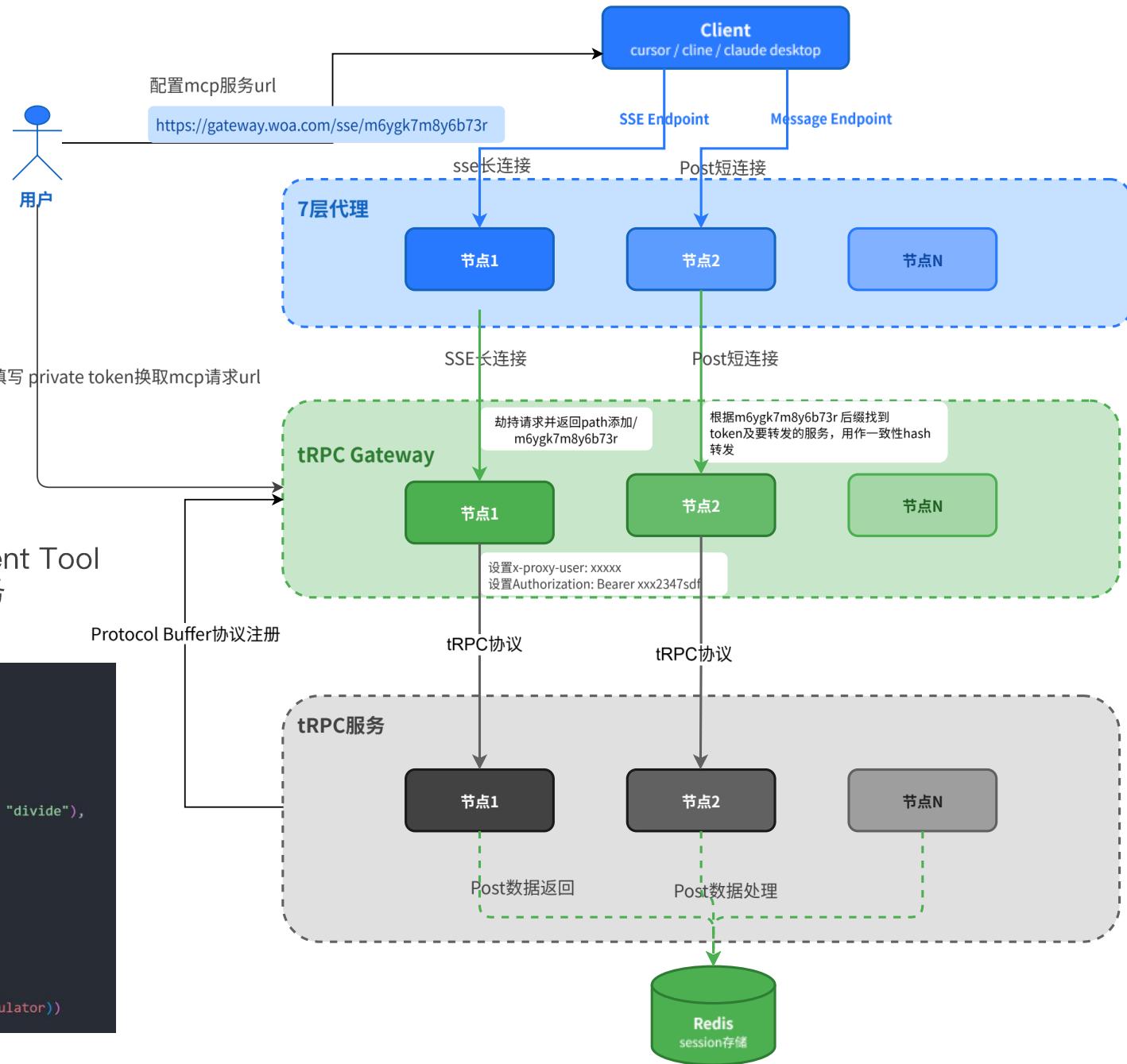
- 自研实现tRPC-MCP-Go框架，结合tRPC SSE协议
- 基于Redis存储，重构节点会话，解决多节点会话丢失问题
- 协议转换：
 - 基于PB注解，一键生成tRPC MCP服务
 - 基于PB注释，解析AST树，一键推断生成tRPC MCP服务/Agent Tool
 - 基于PB协议和MCP Gateway，直接代理tRPC服务为MCP服务

```
// 计算器请求
message CalculatorRequest {
    string operation = 1 [
        (trpc.mcp.field_options).required = true,
        (trpc.mcp.field_options).description = "要执行的数学运算",
        (trpc.mcp.field_options).enum_values = "add",
        (trpc.mcp.field_options).enum_values = "subtract",
        (trpc.mcp.field_options).enum_values = "multiply",
        (trpc.mcp.field_options).enum_values = "divide"
    ];

    double a = 2 [
        (trpc.mcp.field_options).required = true,
        (trpc.mcp.field_options).description = "第一个操作数"
    ];

    double b = 3 [
        (trpc.mcp.field_options).required = true,
        (trpc.mcp.field_options).description = "第二个操作数"
    ];
}
```

```
// 注册计算器工具
mcpServer.AddTool(mcp.NewTool(
    "calculator",
    mcp.WithDescription("执行基本的数学运算"),
    mcp.WithString("operation",
        mcp.Required(),
        mcp.Description("要执行的数学运算"),
        mcp.Enum("add", "subtract", "multiply", "divide"),
    ),
    mcp.WithNumber("a",
        mcp.Required(),
        mcp.Description("第一个操作数"),
    ),
    mcp.WithNumber("b",
        mcp.Required(),
        mcp.Description("第二个操作数"),
    ),
), mcpserver.ToolHandlerFunc(wrapper.handleCalculator))
```



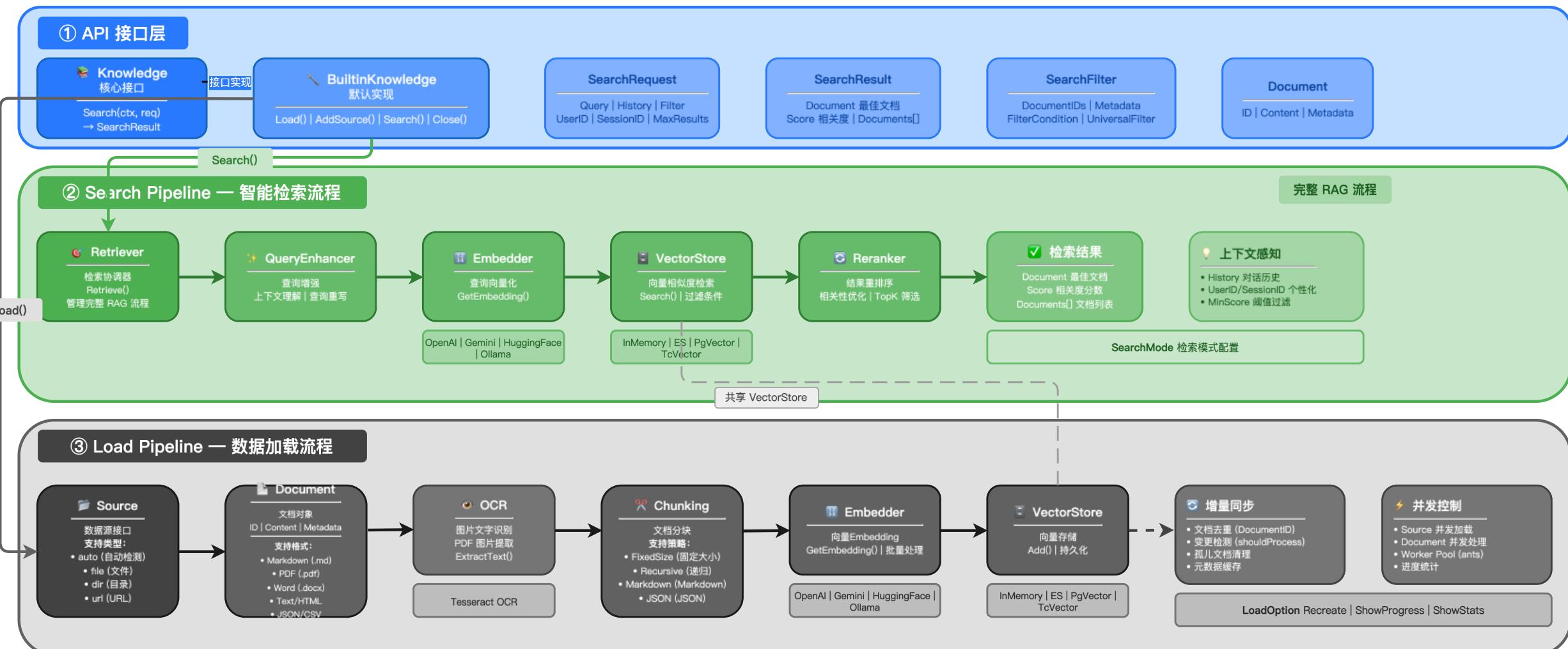
Knowledge

挑战:

- 内容增量同步
- 图片识别
- 大文件内存占用、加载速度慢
- 多轮会话准确性

方案:

- SHA256(Source+URI+内容+分块索引+元数据) 生成确定性ID, 定期对source进行遍历清理, 孤儿文件删除
- 接入Tesseract OCR、多模态模型
- 流式读取、两级并发Source和文档级, 支持远程Embedding
- Search Request加入会话历史用于改写



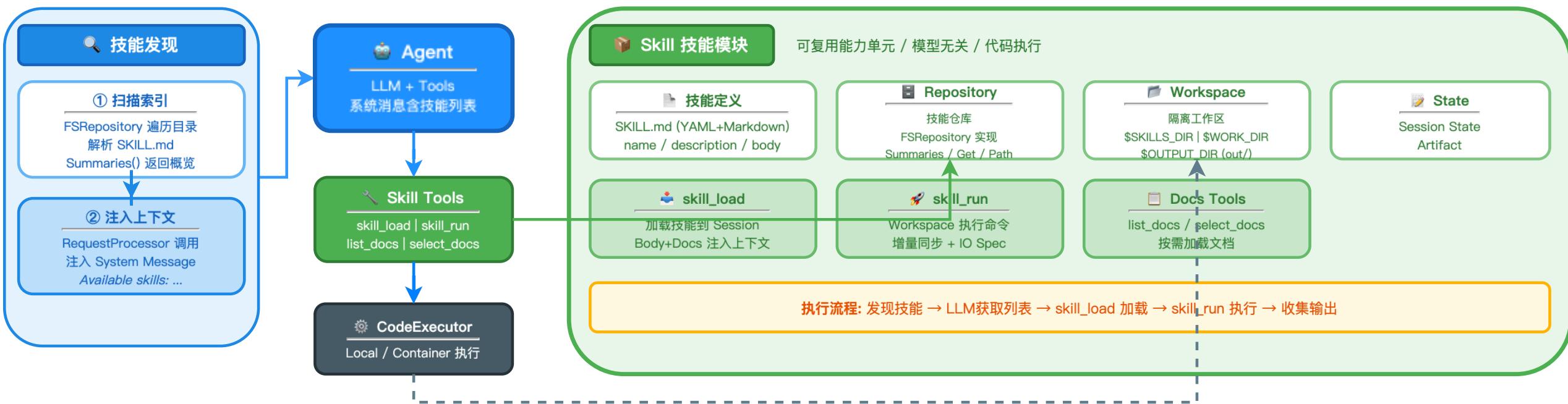
Skills

挑战:

- Claude 模型对Skill原生支持，如何扩展到所有模型的支持，减少Agent上下文开销？
- Skill可移植性、执行安全性问题怎么解决？

方案:

- 概览 (Name/description) 全量注入；正文/文档仅在需要时注入
- 固定目录 (skills/work/inputs/out/runs) + 固定环境变量 (如 WORKSPACE_DIR 等) 保证命令可移植，skill挂载只读防篡改。结合 CodeExecutor、支持强隔离的沙箱环境执行脚本，如Kata Container



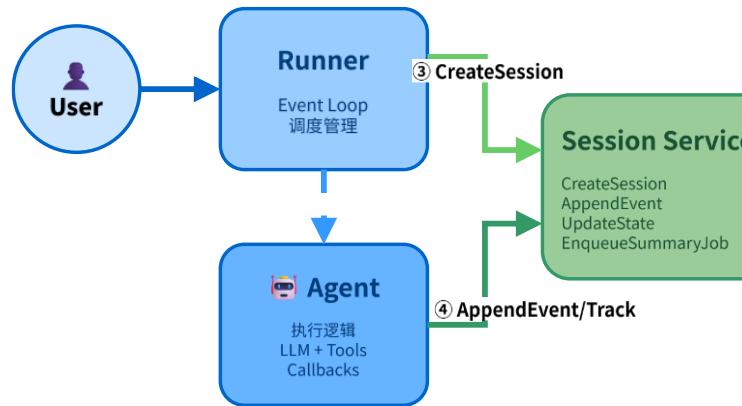
Session

挑战：

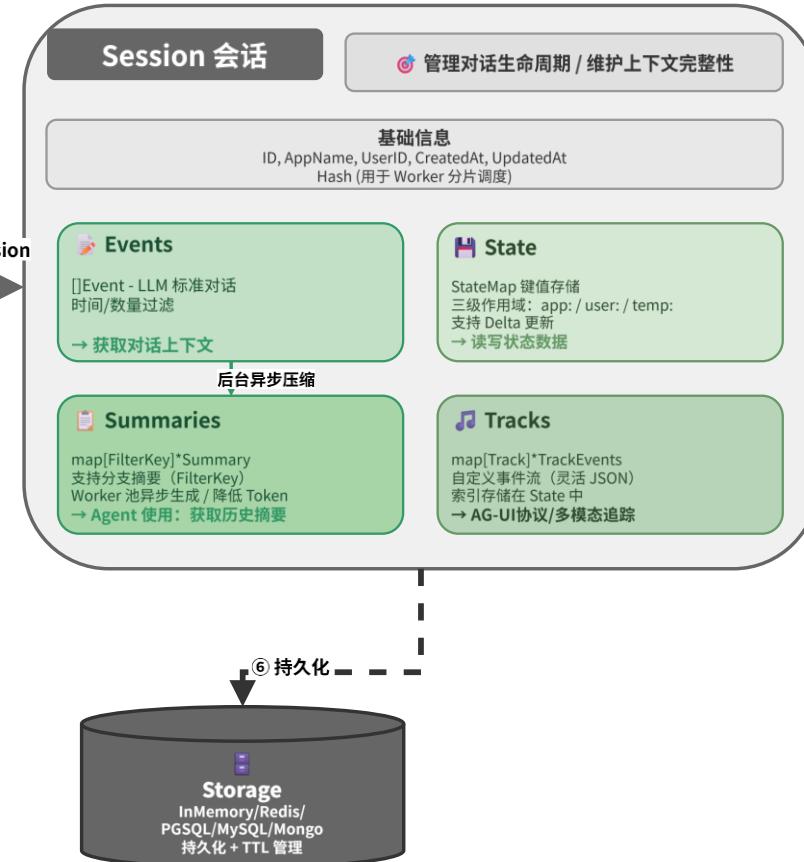
- 会话层级/作用域设计
- 历史会话过长
- 自定义事件流

解决方案：

- Session 支持App、User分层，State根据App、User、Temp前缀区分作用域
- 通过TTL移除历史会话
根据Max Token实现Token自动裁剪
采用Worker pool +Hash路由进行异步通过
Delta 算法summary
- 设置Session Tracks，支持存储自定义事件流，如AG-UI等



Session 架构与 Agent 协作



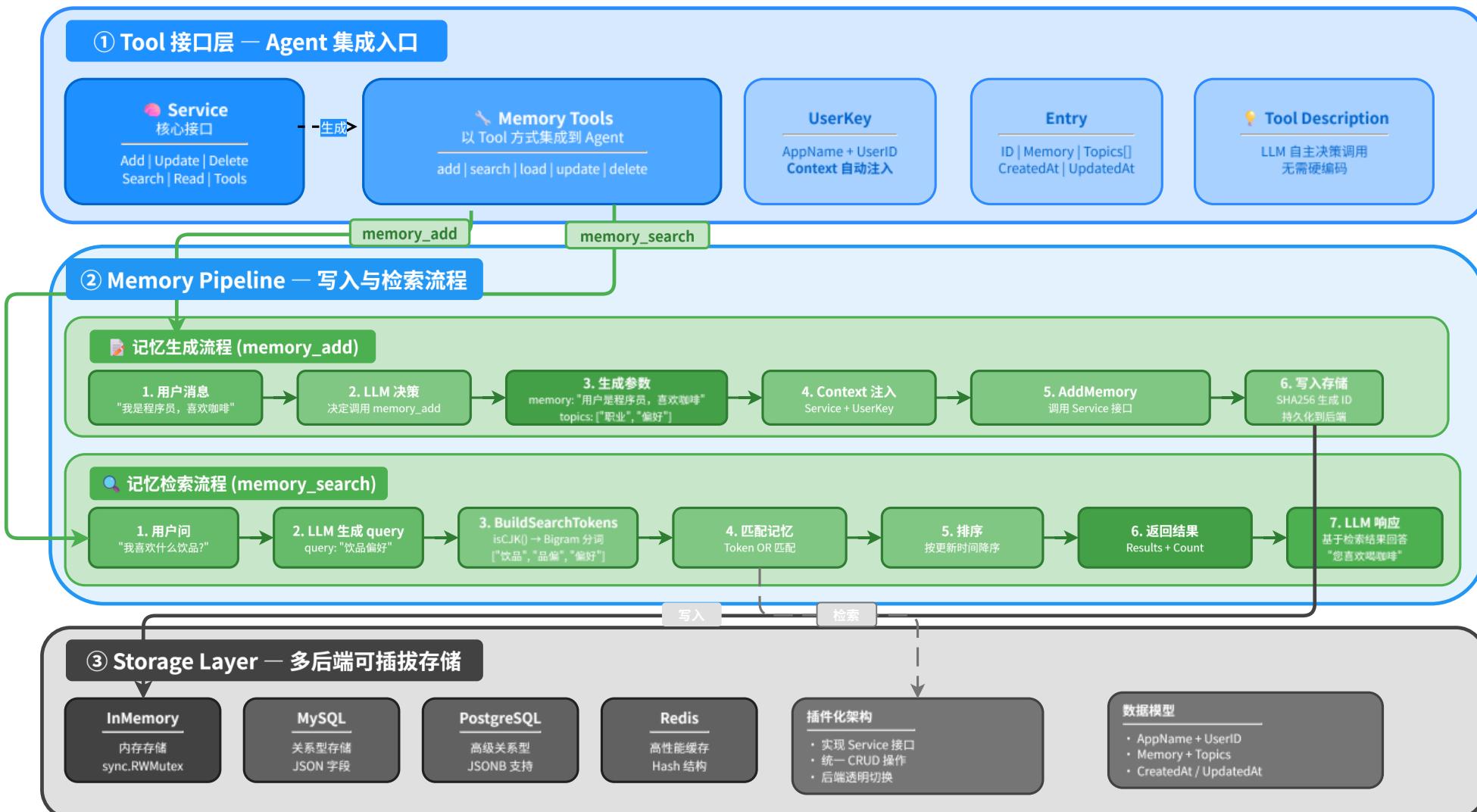
Memory

挑战：

- Memory提取和检索，支持多存储后段
- 中英文混合支持
- 检索效率、准确性

方案：

- 以memory_add、memory_search等Tool方式添加到Agent，Agent决定何时提取/检索记忆，方便插件化扩展MySQL/PGSQL/Redis等
- 自动语言检测，根据是否包含CJK字符，中文采用Bigram分词
- 采用Topics对记忆进行检索分类，提升检索的效率



AG-UI/A2A

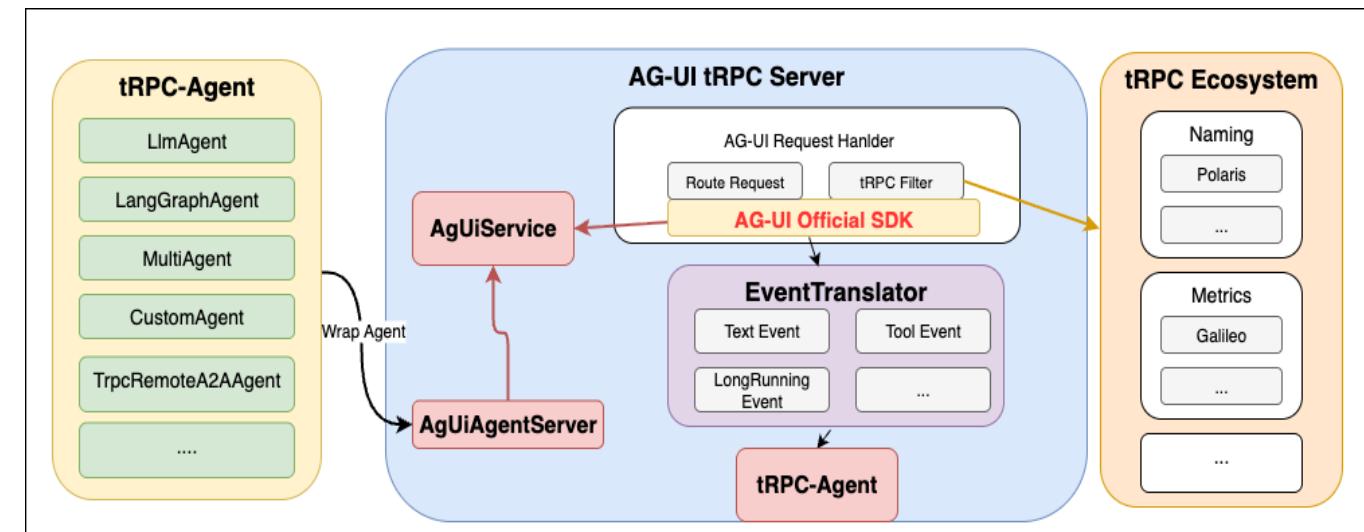
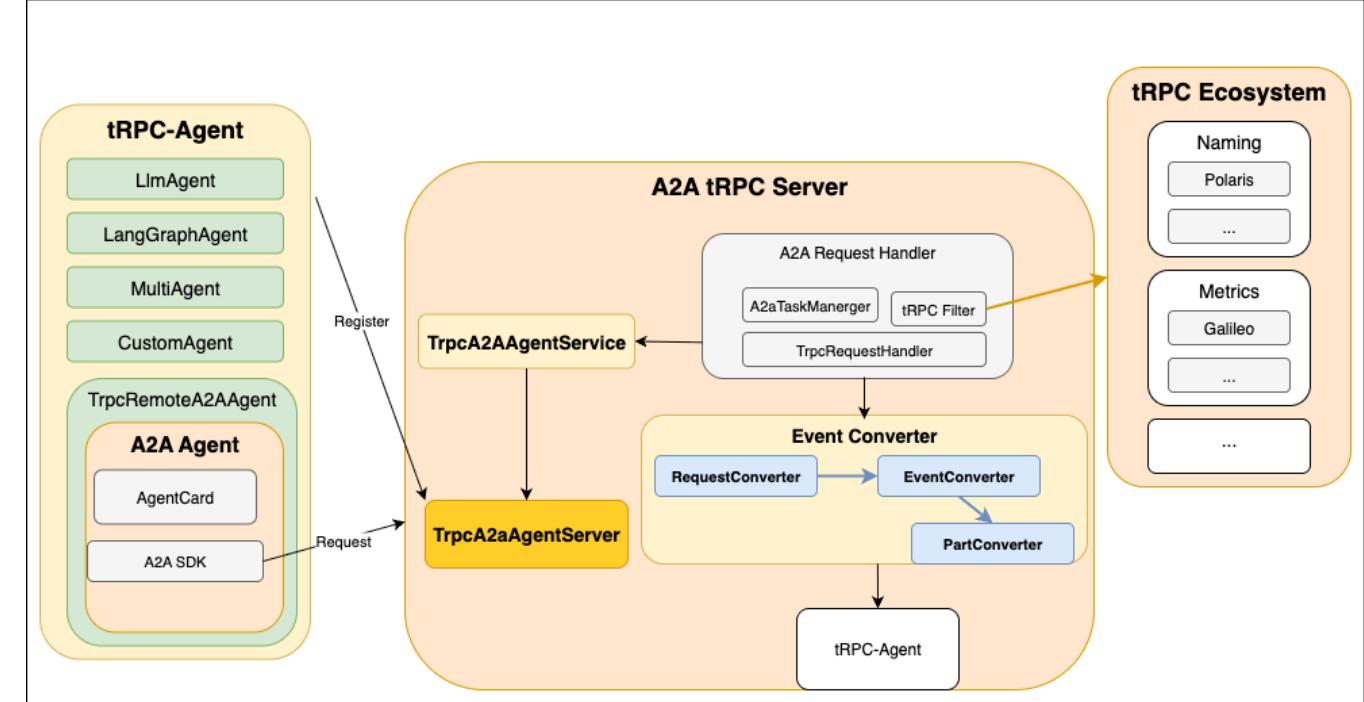
挑战：

- 早期A2A/AG-UI 没有Golang 官方SDK
- A2A 服务远程调用复杂
- A2A、AG-UI事件和Agent事件转换
- 和tRPC框架的结合

方案：

- 实现tRPC-A2A-Go并对外开源，为最早的开源A2A Go SDK, CNCF项目在使用；
AG-UI从第一个Go MR合入就开始共建，贡献多个功能和修复
如：工具执行结果事件、用户上报执行进度、SSE 新增Flush逻辑、完善事件序列化方法等

- 实现了A2AAgent, Agent和本地调用Agent一致，只需传入 Agent Card
- 实现Event Converter/Translator, 将Agent内部Event转化成 A2A Message/AG-UI事件
- 对外基于tRPC SSE协议提供服务



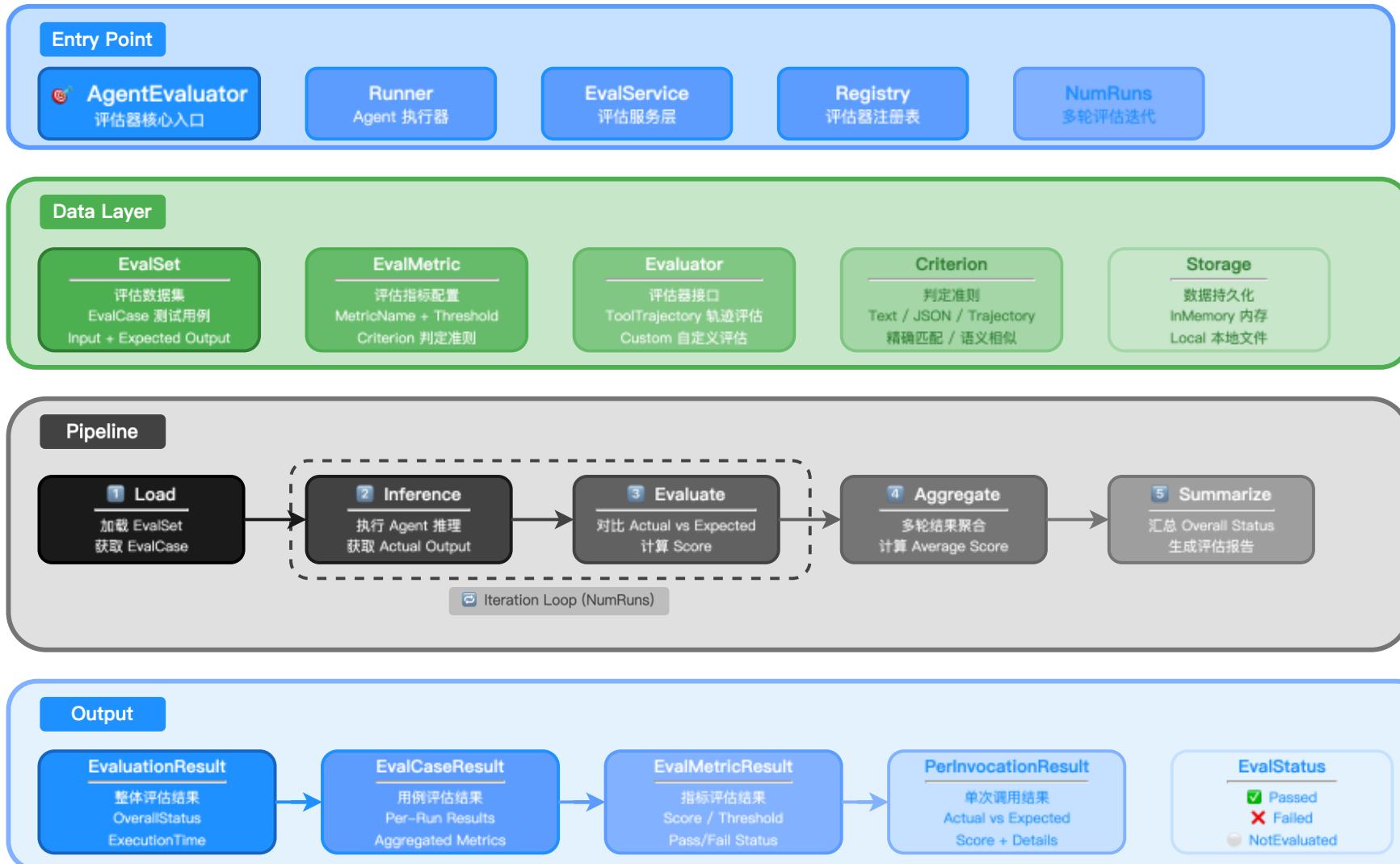
Evaluation

挑战:

- 大模型输出的不确定性
- 无GroundTruth
- 评测数据集构建复杂

方案:

- 工具轨迹 (ToolTrajectory) 评测, 多轮评估支持
- LLM-as-Judge (Faithful、Correctness)
- 和可观测系统结合, 把trace数据转换成评估集, 使用不同可观测系统的评估器进行评估和结果数据上报



04 tRPC-Agent业务实践

腾讯元宝

结合tRPC-Agent的GraphAgent串联多个子LLM Agent，完成用户写小说的场景





【腾讯体育app】团队基于tRPC-Agent搭建了覆盖比赛前/中/后全链路的一站式AI观赛助手“体育大聪明”，支持在搜索、社区评论区、直播间弹幕区、底tab等多个场景与“体育大聪明”对话，回答用户各种体育观赛问题。同时团队还进一步使用Agent能力拓展实现了AI中文解说、AI比赛战报等特色应用。

The image displays five screenshots of the Tencent Sports app illustrating its AI integration:

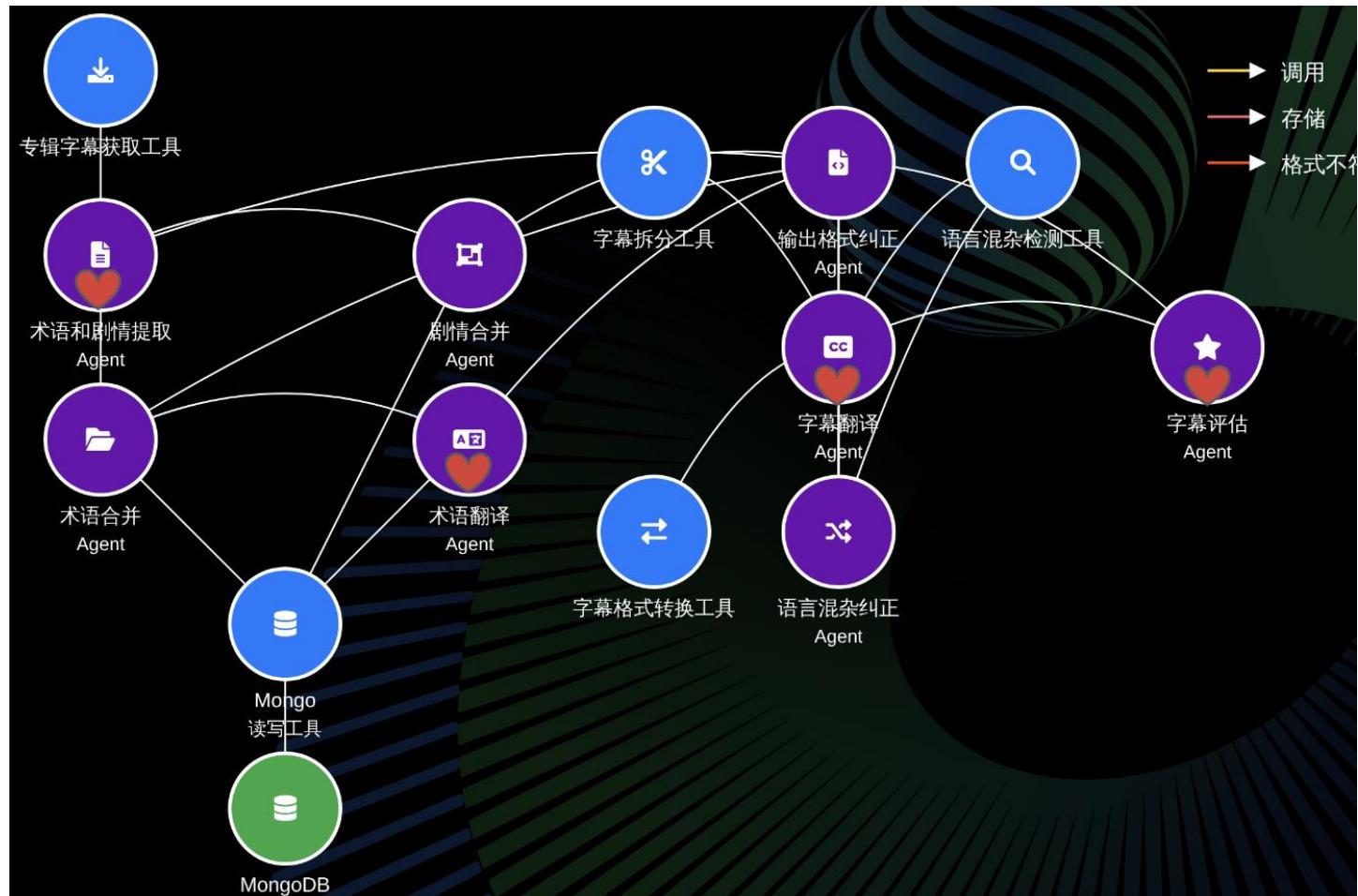
- 体育搜索**: Shows a search results page for NBA东西部排名, listing teams by胜 (Wins), 负 (Losses), and 胜率 (Win Rate). A sidebar for "体育大聪明" provides AI-generated insights.
- 底tab入口**: Shows the bottom tab bar with icons for 首页 (Home), 赛程 (Schedule), 会员 (Membership), and 我的 (My Account). An "问AI" button is also present.
- AI战报**: Shows a news article about LeBron James' career records, with a large image of him and a "球员数据" (Player Data) section.
- 社区互动**: Shows a comment section for a dynamic post, displaying user interactions and AI-generated responses.
- 直播间互动**: Shows a basketball game in progress with a live chat overlay from the "开拓者" (Portland Trail Blazers) team.
- AI中文解说**: Shows an AI-generated Chinese broadcast interface with options for different accents (Wang Zhaofeng, Wenrui Shuai, Cantonese, Sichuanese, Northeastern), player profiles, and a "取消" (Cancel) button.

腾讯视频WeTv

WeTv 字幕翻译：基于 trpc-agent 框架搭建了下面四个 llmagent：

- 剧情提取 agent
- 术语提取 agent
- 术语翻译 agent
- 字幕翻译 agent

首先会使用剧情提取和术语提示 agent 处理专辑的全部字幕，输出剧情和术语表，然后用术语翻译 agent 将术语表翻译为海外 12 种语言，最后将字幕、剧情和术语一并输入到字幕翻译 agent 来得到字幕翻译的最终结果。



腾讯业务落地和开源

腾讯有120+部门在使用和试用tRPC-Agent生态

主要业务场景：元宝、体育、视频、QQ、文档、新闻、音乐、AMS广告、IMA、微信支付、Fit财付通，腾讯云、先锋游戏等业务的都有场景应用



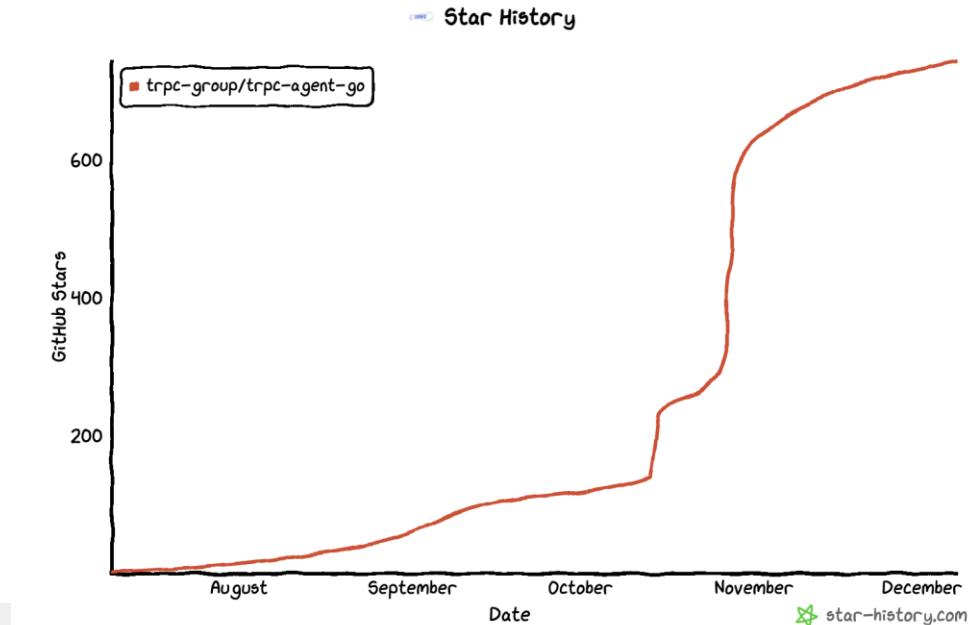
.....

开源地址：<https://github.com/trpc-group/trpc-agent-go>

tRPC-Agent-Go

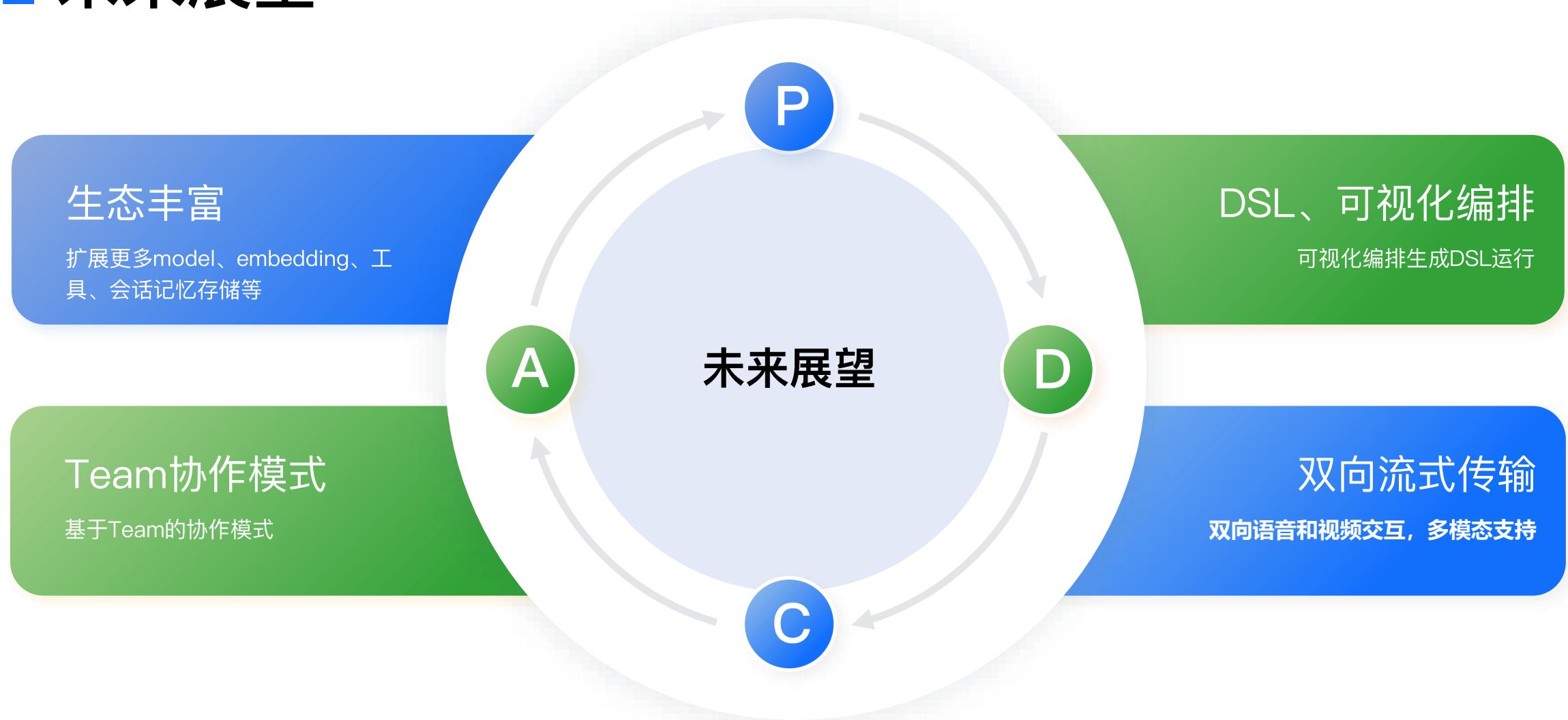
[reference](#) [go report](#) [A+](#) [license Apache-2.0](#) [release v0.7.0](#) [Pull Request Check](#) [passing](#) [codecov 88%](#) [Docs](#) [Website](#)

A powerful Go framework for building intelligent agent systems that transforms how you create AI applications. Build autonomous agents that think, remember, collaborate, and act with unprecedented ease.



06 tRPC-Agent未来展望

未来展望



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

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

1000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

上海

1200人

THANKS

探索 AI 应用边界

Explore the limits of AI applications

AiCon

全球人工智能开发与应用大会

