可扩展的智能体（Agent）系统框架：基于大语言模型的智能决策平台设计

汇报人：李佑天、杜沛生、梁研锋

目录

CONTENT

项目背景

01

项目背景

1. 人工智能的应用是大势所需

- 全球范围内，多国将智能体技术纳入战略规划。
- 中国“十四五”规划中明确提及“推动人工智能与实体经济深度融合”。
- 智能体作为具身智能、自主系统的核心载体，在智能制造、智慧城市等领域获得政策支持。

2 特向训练或多端融合

- 强化学习： DeepMind的AlphaGo 通过不断博弈达到围棋领域领先。
- 融合训练：特斯拉的自动驾驶视觉系统融合多摄像头数据实现辅助驾驶。

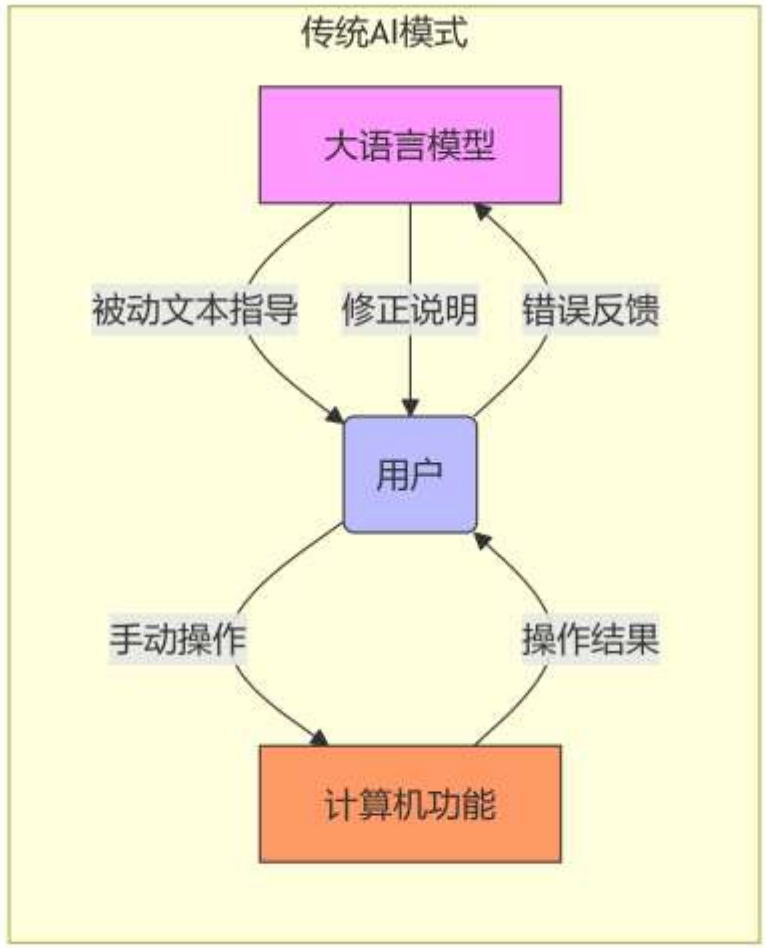
这些应用方向都偏向针对性，少有通用化的解决方案。

项目背景

	智能体（Agent，如 YAA）	**普通 AI**（如传统聊天机器人）
自主性：	智能体能**主动感知环境并做出决策**，而非被动响应用户输入	被动响应，需用户触发（如问答）
动态环境适应能力：	能在**不确定环境中持续学习并调整策略**，适应规则变化或新场景。	无实时环境交互能力
记忆与学习：	具备**持续记忆和历史交互学习能力**，提供多元化、个性化的服务。	通常无记忆或短期会话记忆
多工具协同：	可**调用多种工具/API**，完成跨平台、多步骤任务。	功能单一

项目背景

案例：用户想要配置Java环境



普通AI局限：

被动文本指导：仅能输出步骤说明（例如“访问Oracle官网下载JDK，然后设置环境变量……”），无法执行实际安装操作。无法根据用户系统环境（如Windows/macOS差异、权限状态）动态调整指导方案。无执行纠错能力：若用户操作错误（如路径填写错误），无法主动干预修复。

项目背景

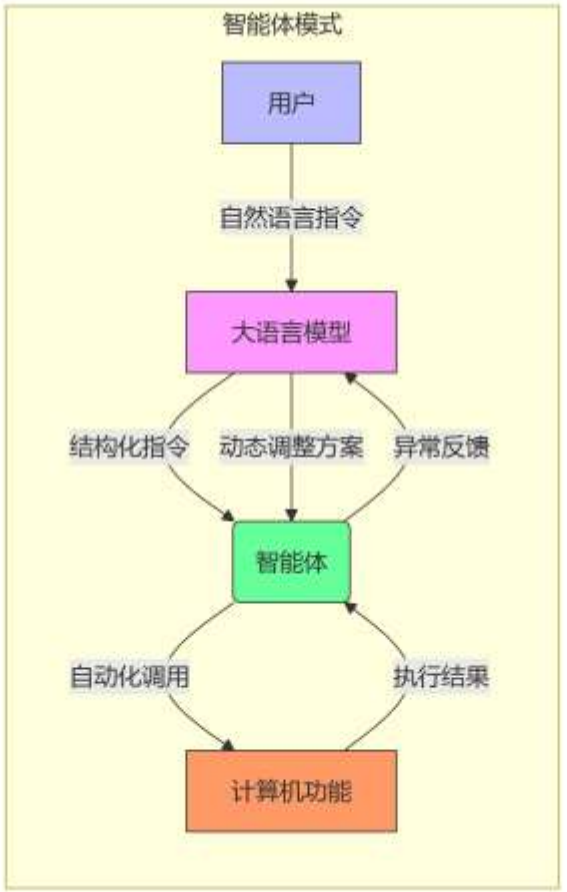
案例：用户想要配置Java环境

智能体能力：

自动化环境检测与配置：通过调用系统工具（如PowerShell/Shell）直接检测操作系统类型、现有Java版本，并动态下载适配的JDK。

权限调用与静默安装：获取管理员权限后自动完成JDK安装、配置环境变量（如JAVA_HOME、PATH），无需用户手动操作。

异常处理：若安装失败（如网络中断），主动触发重试机制或提供修复建议（如清理残留文件）。



作品介绍

Project
Introduction

02

作品介绍

YAA 意思是：又一个智能体（Yet Another Agent），是一个能分析理解自然语言指示，自动创建、规划、执行、验证任务的智能体。

作品介绍

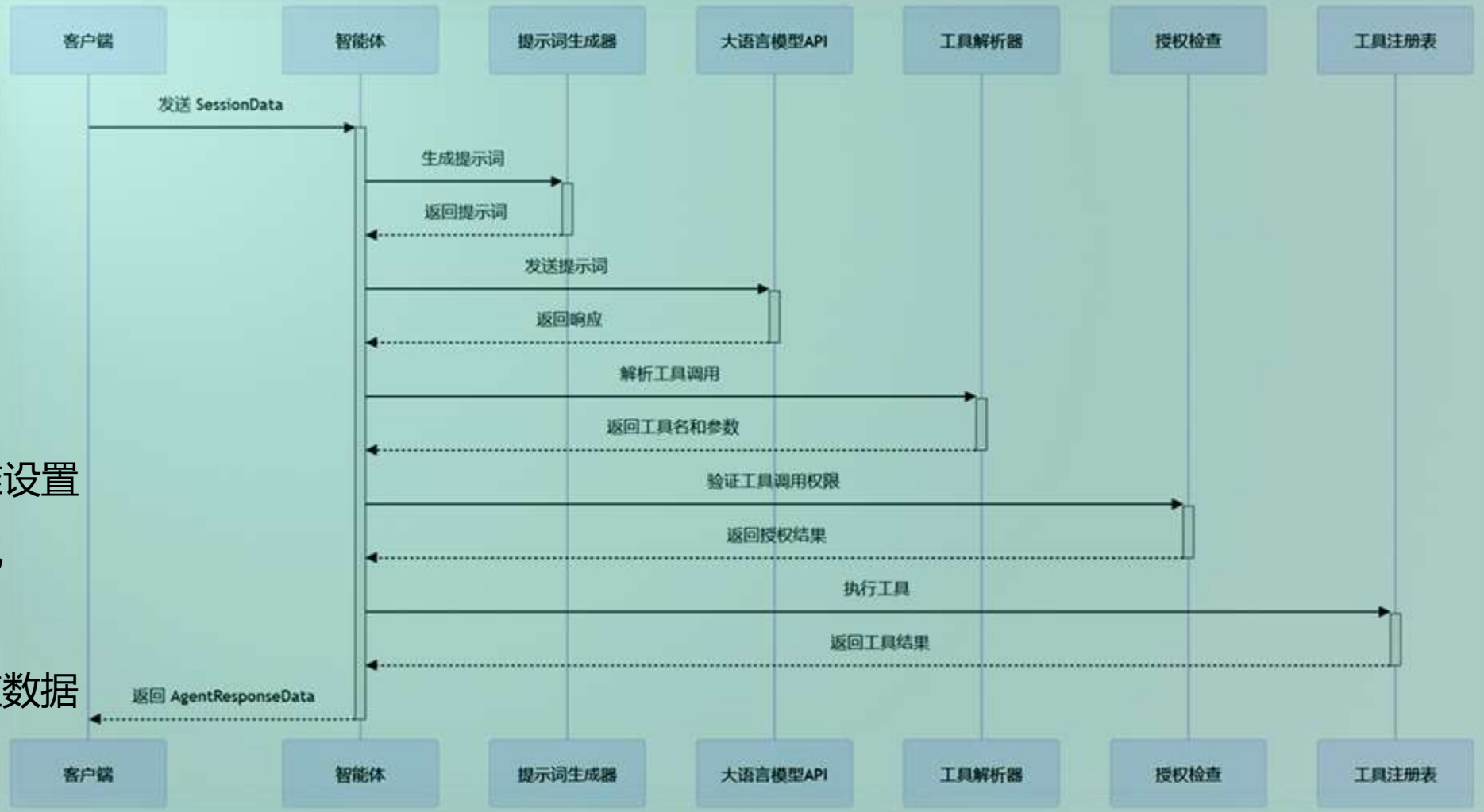
创新点、技术指标及设计目的:

高兼容性 智能体服务端采用 Rust 实现，支持 WASM（WebAssembly）和 WASI 运行时环境。单次编译后可部署在：浏览器运行时、用户设备、云沙盒隔离执行环境。

高灵活性 通过提示词提取实现大语言模型的工具调用，无需额外训练，即可快速适配不同领域任务。智能体支持动态加载插件，可根据需求灵活扩展功能模块，实现开箱即用的多场景支持。在极端条件下还可利用大语言模型的代码能力给自身编写针对性的插件。

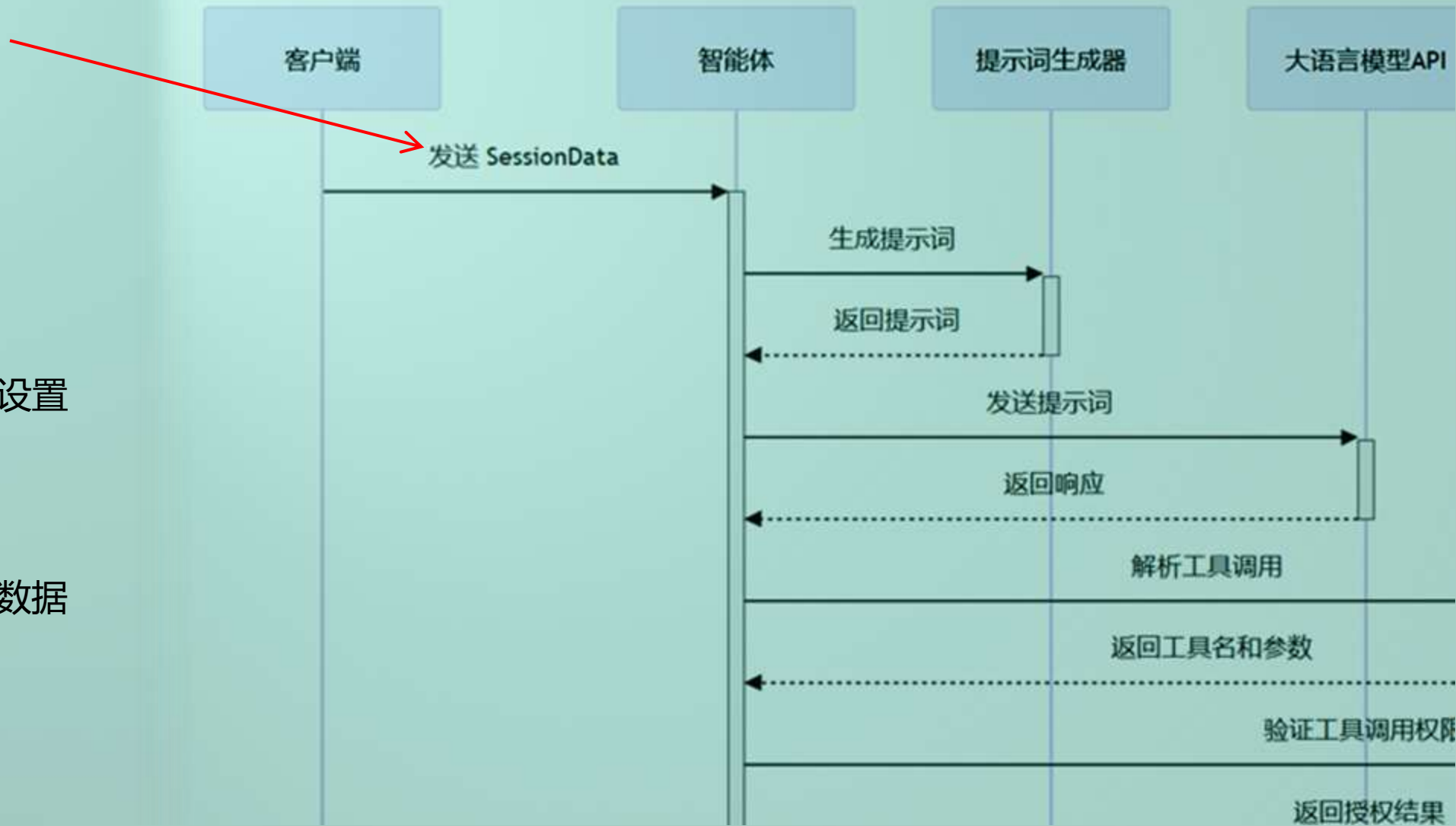
作品介绍 核心工作流程

- 接收会话数据
- 检查并补全会话配置
- 生成提示词
- 调用大语言模型 API
- 解析工具调用
- 检查工具授权
  - 检查会话配置自动批准设置
  - 检查用户消息授权标记
- 执行工具
- 判断是否结束，生成响应数据



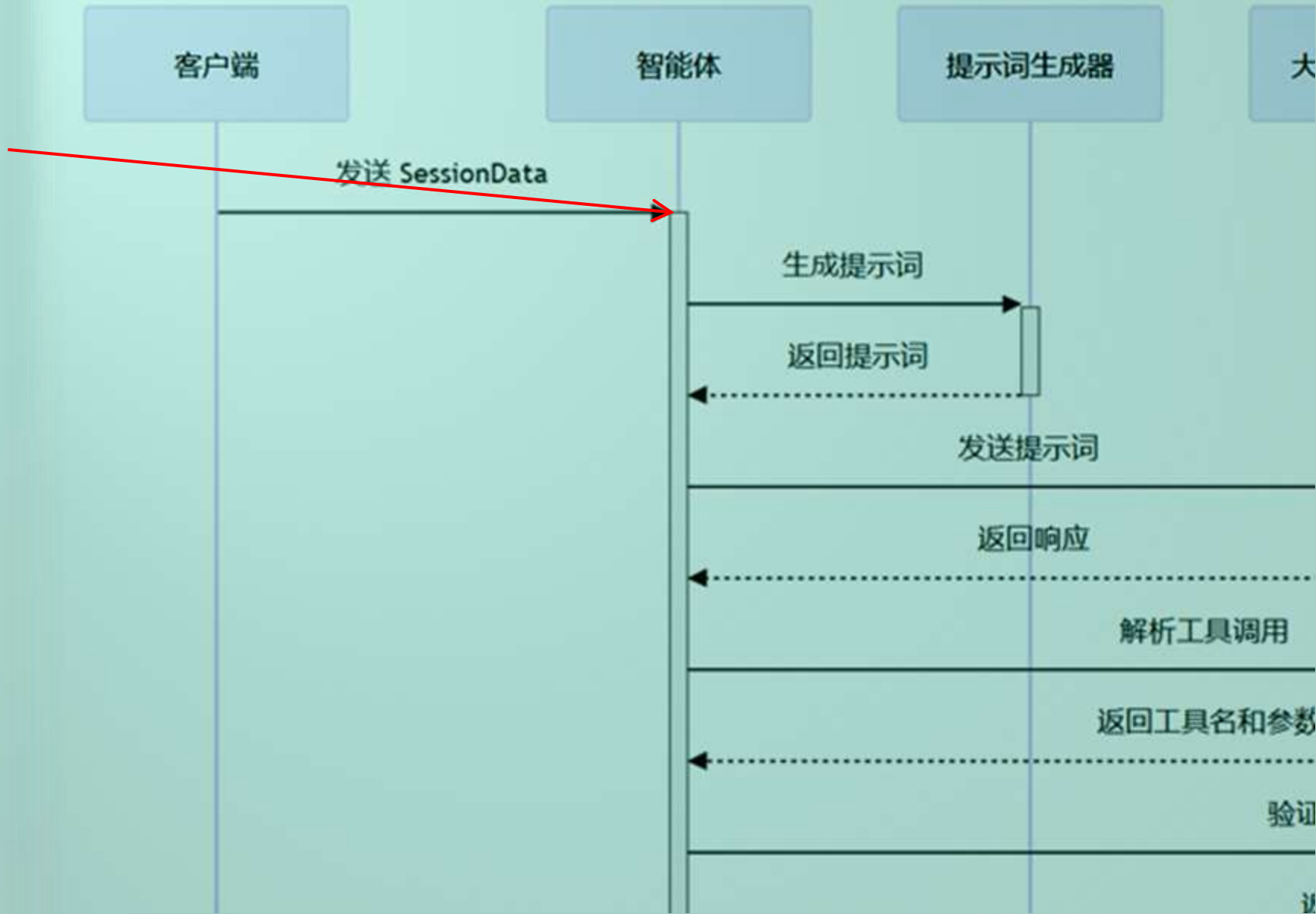
作品介绍 核心工作流程

- 接收会话数据
- 检查并补全会话配置
- 生成提示词
- 调用大语言模型 API
- 解析工具调用
- 检查工具授权
  - 检查会话配置自动批准设置
  - 检查用户消息授权标记
- 执行工具
- 判断是否结束，生成响应数据



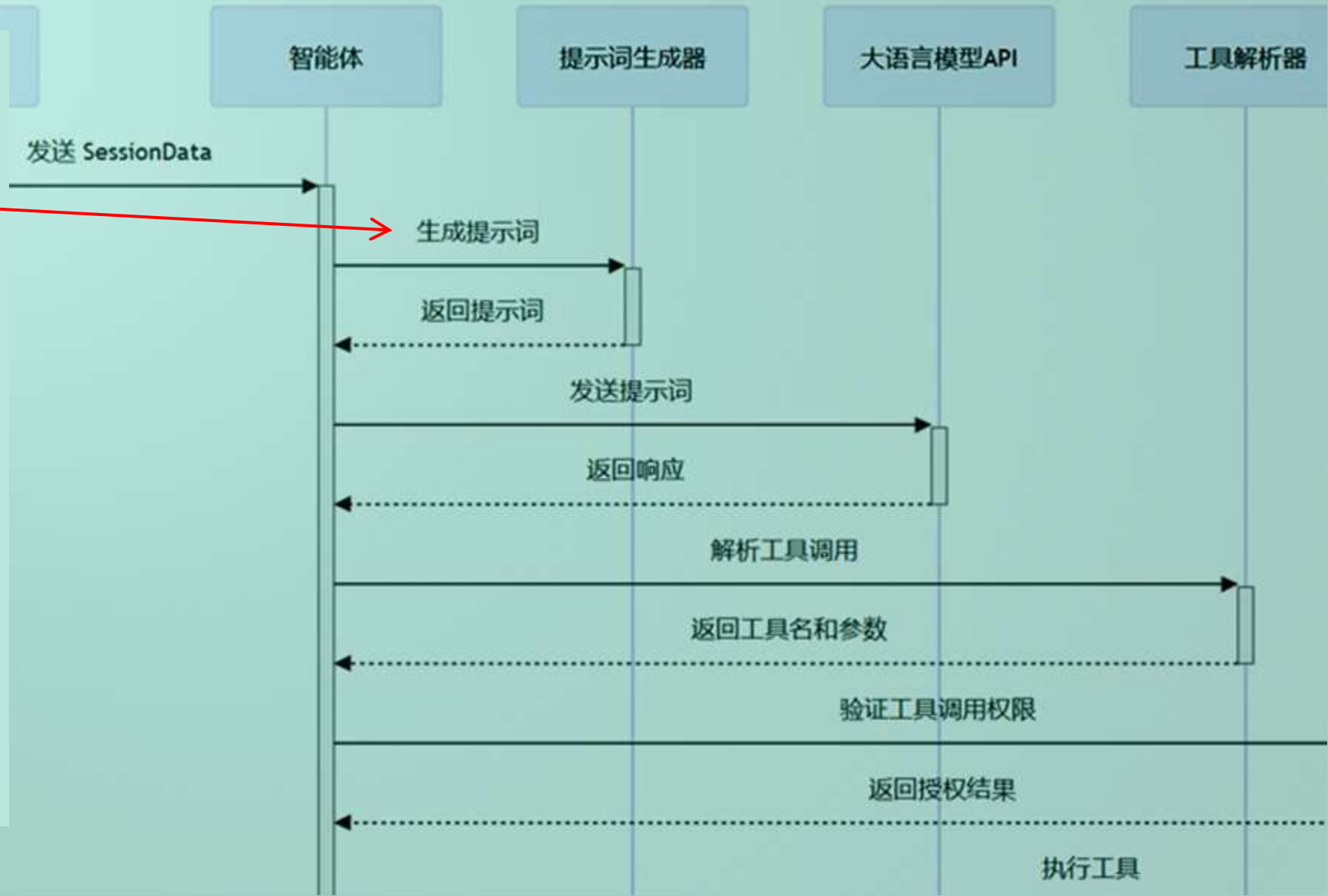
作品介绍 核心工作流程

- 接收会话数据
- 检查并补全会话配置
- 生成提示词
- 调用大语言模型 API
- 解析工具调用
- 检查工具授权
  - 检查会话配置自动批准设置
  - 检查用户消息授权标记
- 执行工具
- 判断是否结束，生成响应数据



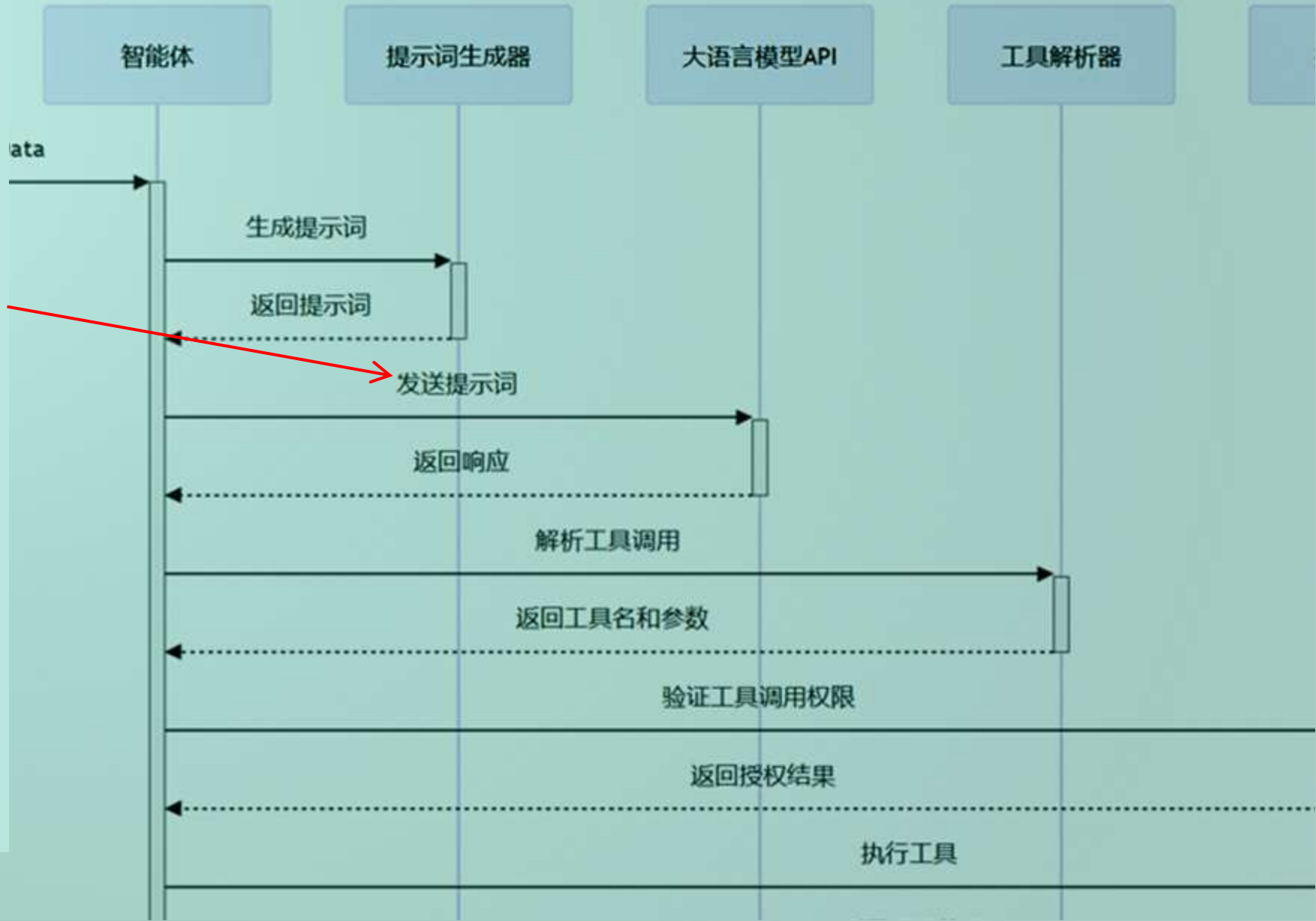
作品介绍 核心工作流程

- 接收会话数据
- 检查并补全会话配置
- **生成提示词**
- 调用大语言模型 API
- 解析工具调用
- 检查工具授权
  - 检查会话配置自动批准设置
  - 检查用户消息授权标记
- 执行工具
- 判断是否结束，生成响应数据



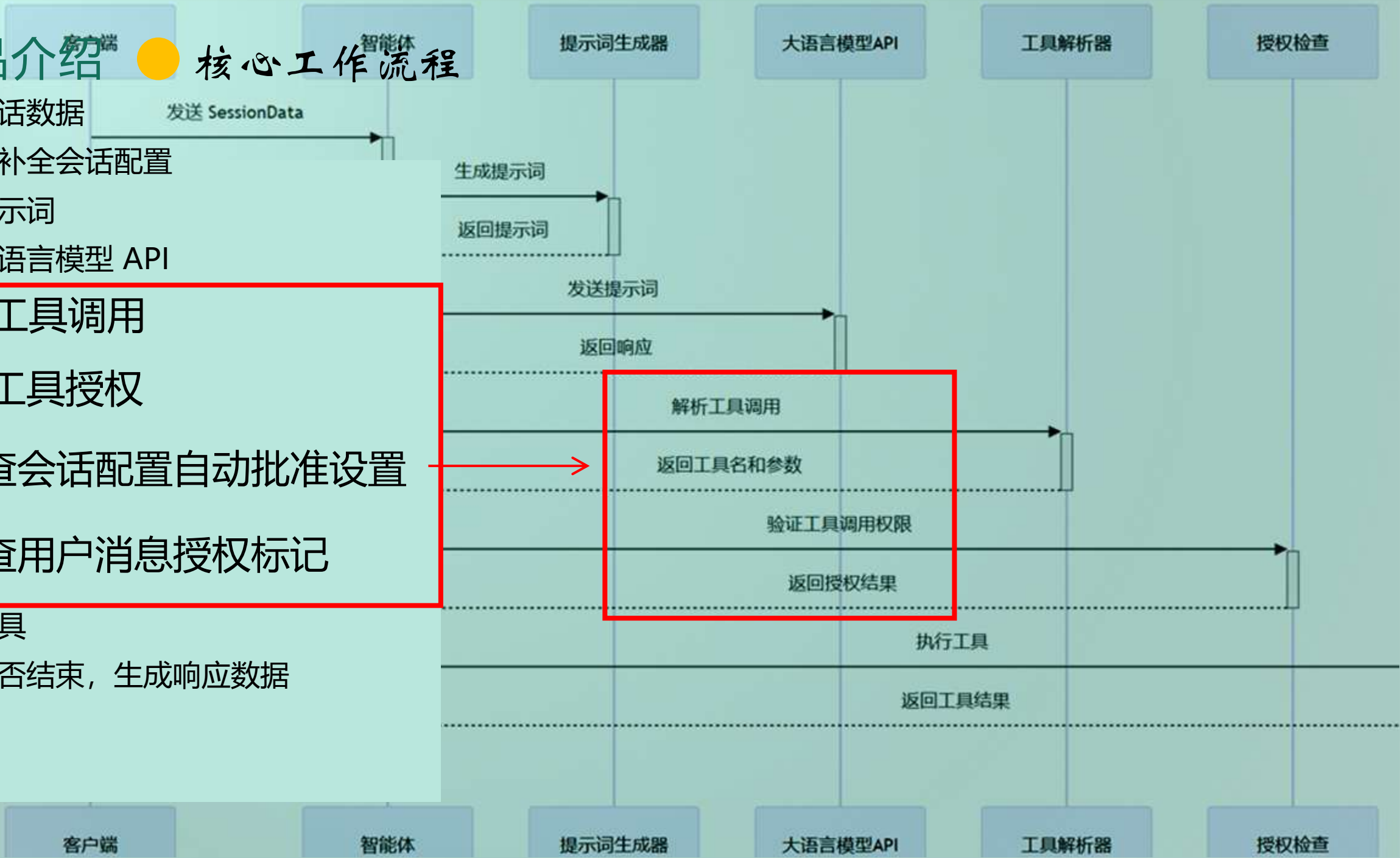
作品介绍 核心工作流程

- 接收会话数据
- 检查并补全会话配置
- 生成提示词
- 调用大语言模型 API
- 解析工具调用
- 检查工具授权
  - 检查会话配置自动批准设置
  - 检查用户消息授权标记
- 执行工具
- 判断是否结束，生成响应数据



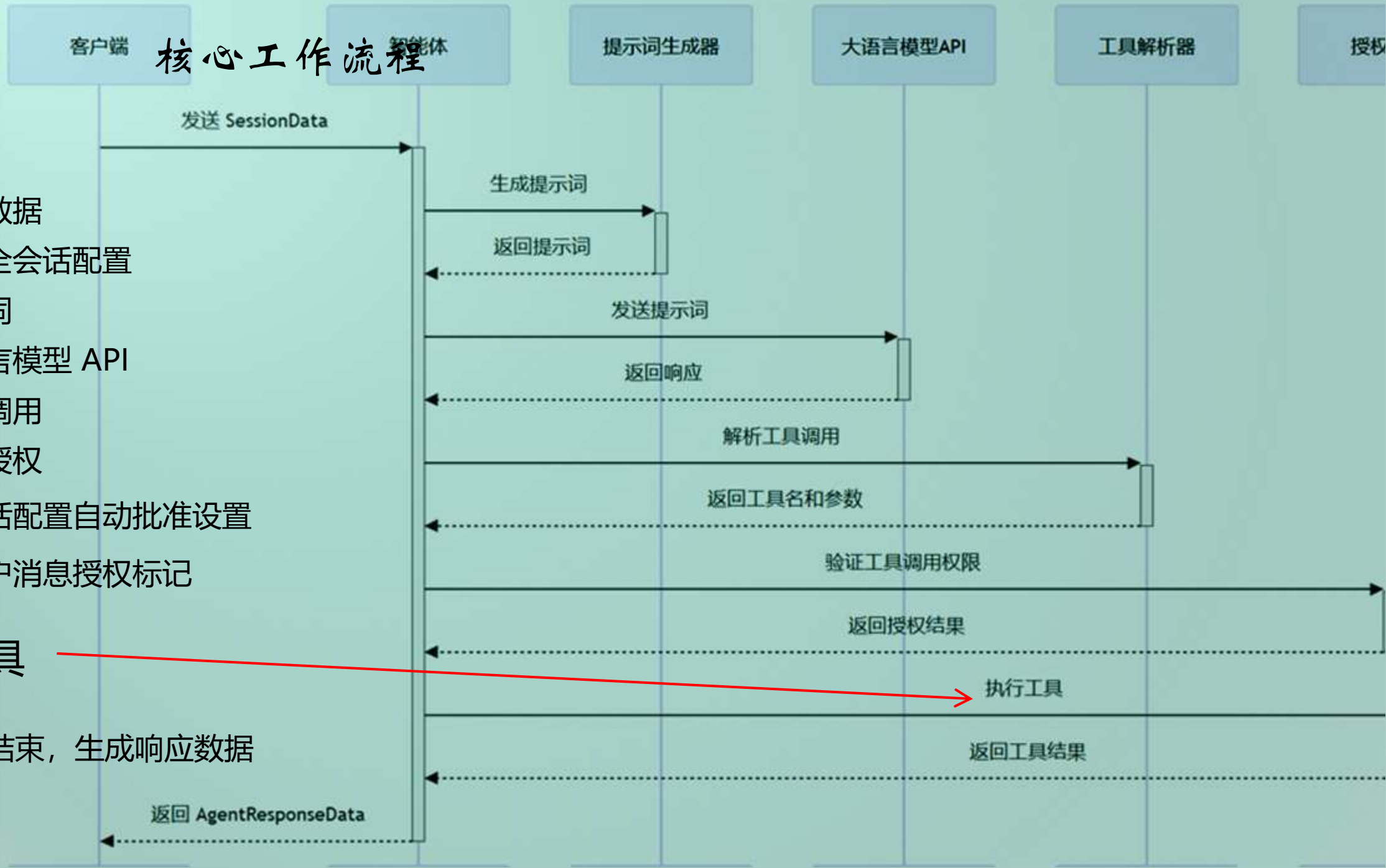
作品介绍 核心工作流程

- 接收会话数据
- 检查并补全会话配置
- 生成提示词
- 调用大语言模型 API
- 解析工具调用
- 检查工具授权
  - 检查会话配置自动批准设置
  - 检查用户消息授权标记
- 执行工具
- 判断是否结束，生成响应数据



作品介 核心工作流程

- 接收会话数据
- 检查并补全会话配置
- 生成提示词
- 调用大语言模型 API
- 解析工具调用
- 检查工具授权
  - 检查会话配置自动批准设置
  - 检查用户消息授权标记
- **执行工具**
- 判断是否结束，生成响应数据



作品介绍 核心工作流程

- 接收会话数据
- 检查并补全会话配置
- 生成提示词
- 调用大语言模型 API
- 解析工具调用
- 检查工具授权
  - 检查会话配置自动批准设置
  - 检查用户消息授权标记
- 执行工具
- 判断是否结束，生成响应数据

具体实现

Concrete Realization

核心功能介绍

会话数据管理

用于记录每次会话的数据，包含唯一标识符、标题、会话开始时间、会话状态、消息内容以及配置数据

工具调用框架

基础包括再度思考工具与完成回话工具，允许开发者开发更多工具，扩展智能体能力

提示词生成

通过系统调用提示词生成器，引导智能体输出规范化并赋予其调用工具能力。

大语言模型 API集成

开放自定义API接口，可接入Deepseek、通义千问等多款热门大语言模型

作品介绍

核心流程分析（用户智能体交互）

向服务端发送响应消息并弹出用户消息。

二.yaa 生成根据系统设置生成初始提示词。
三.会话数据传递给大模型。
四.如果用户未为此工具设置自动授权，会话数据传回用户进行授权。

六.会话数据传递给工具调用器。并执行。 重复步骤三、四、五，直到大模型认为任务已完成

工具执行完成发送到客户端并输出最终结果

一.用户输入命令。

作品介绍 功能模块介绍

保存用户自定义设置

展开与折叠侧边栏按钮

重载页面按钮

可建立多个独立会话，支持独立会话的删除功能

YAA设置面板展开，可自定义更换模型

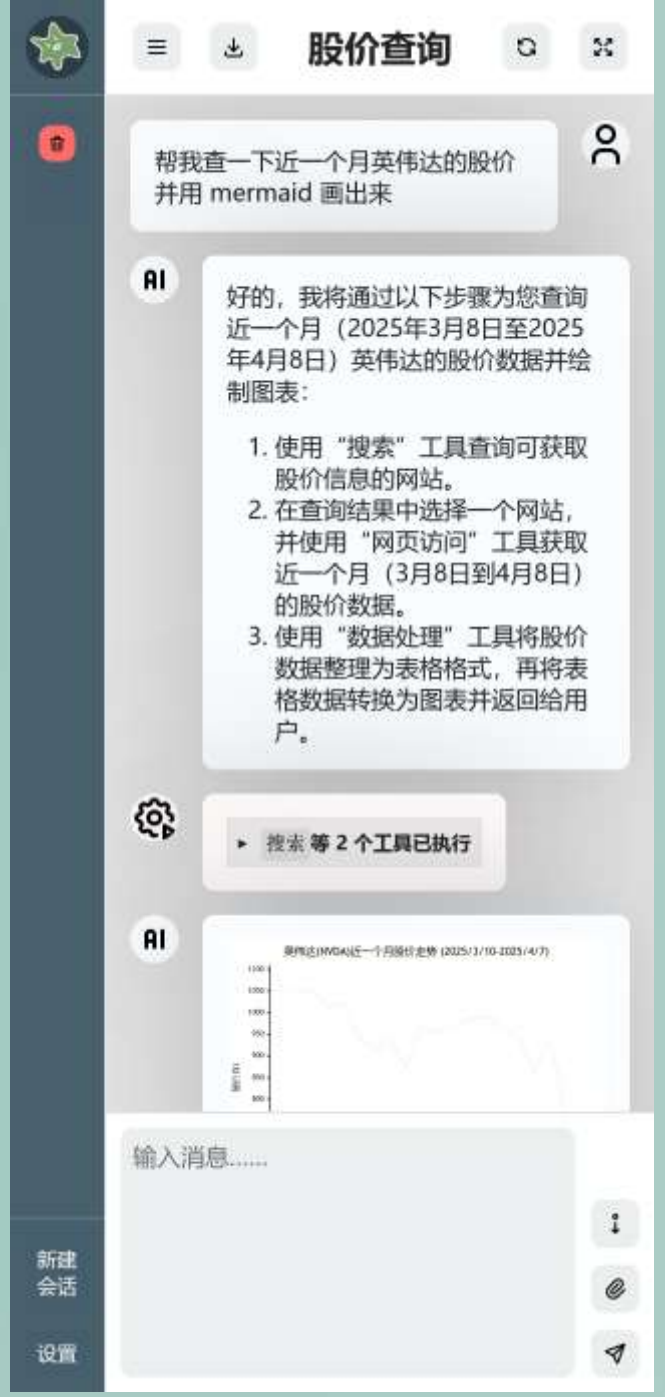
可扩大缩小整个页面

新建会话按钮

回到底部按钮

点击可展开设置界面

作品介绍



响应式设计，兼容多端页面大小

应用前景

智能体赋能新型工业化，为企业起步快速建立工作流。

03

应用前景

Application & Prospect

易用

高效

安全

可扩展

对话即工作

基于大语言模型服务的对话式工作流程。自然交互，低使用门槛。智能体的上下文感知特性使得用户的意图容易被理解，无需用户重复输入。

过程皆自动

通过与智能体交互直接生成工作流，让智能体自动串联多步骤工作流程，减少人工干预，提高AI输出转化为实际工作成果的效率。

操作需授权

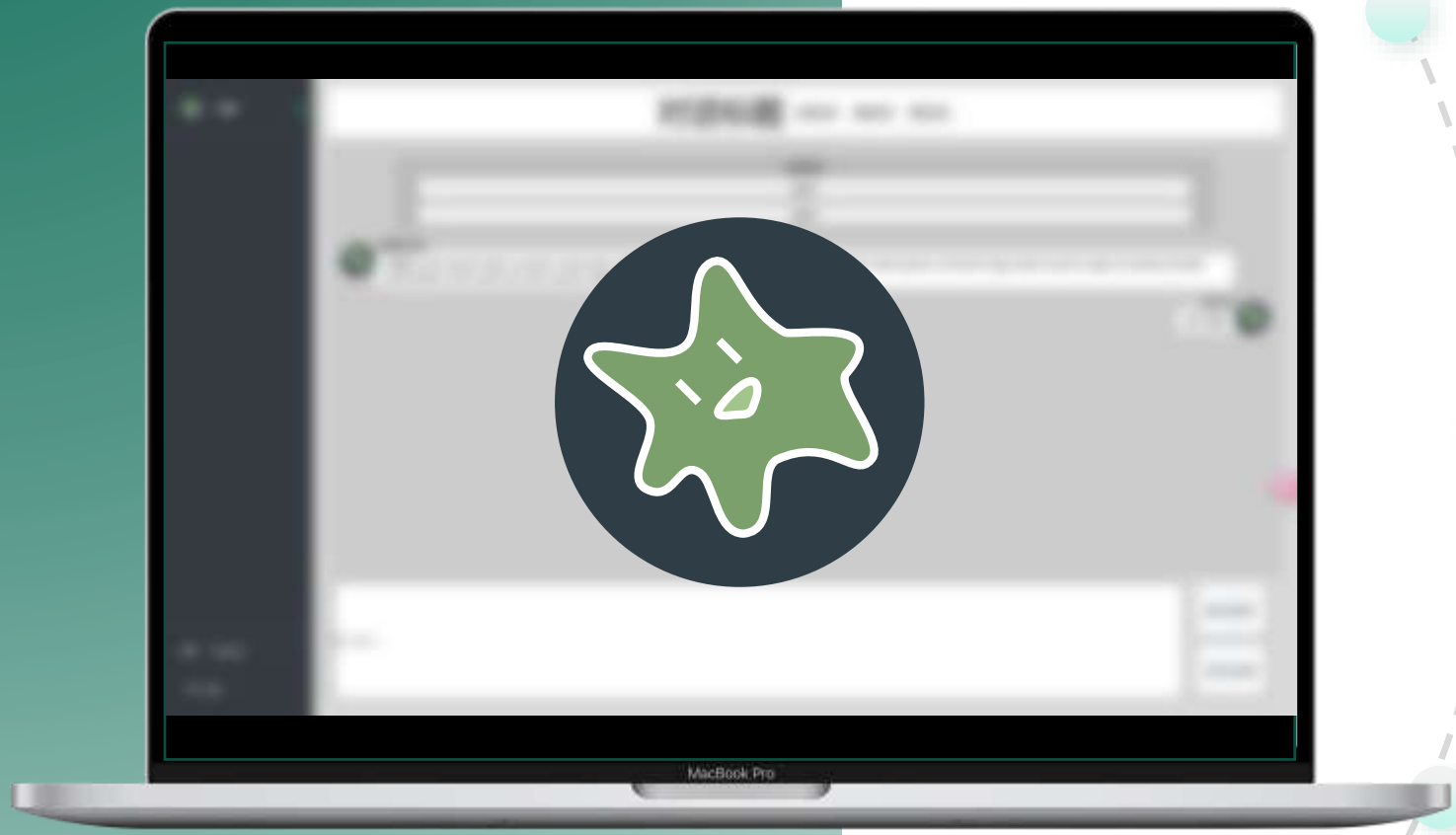
智能体的所有敏感操作均需要经过人工授权，避免智能体越权进行非法操作。用户亦可为提升工作效率，为智能体开启部分乃至完全的自动授权。

扩展随需求

支持MCP协议，允许开发者为智能体开发新的工具，赋予智能体更多能力。可自定义大语言模型API，获取来自不同模型的输出结果。

应用前景

Application & Prospect



01 个人助手

提高日常事务处理效率，节省时间

个人关键事务提醒，减少错误发生率

02 工业自动化

实时控制工业生产线，提高生产效率

动态调整工艺参数，提高产品良率

03 企业决策

即时获取数据分析报告，提高战略决策速度

提高风险预测准确性，减少决策失误率

未来工作计划

NEXT WORK PLAN

•多模态支持

扩展自然语言理解至多模态（图像、音频、视频等），结合视觉模型或语音识别插件，实现更复杂的任务处理（如分析图表、生成语音反馈等）。

•动态代码生成优化

提升智能体自编程插件的可靠性和安全性，例如通过沙盒环境限制、静态代码分析或运行时监控，确保生成的插件代码无漏洞且符合预期。

•任务流自动化增强

引入更复杂的任务规划机制（如依赖管理、条件分支、循环执行），支持长期运行的跨平台工作流（例如自动化数据分析报告生成）。

未来工作计划

NEXT WORK PLAN

“让智能体赋能新时代工业化建设。”

感谢观看

汇报人：李佑天、杜沛生、梁研锋

2025年4月10日