

大模型智能体技术

▲ 主讲： 郭源

▲ 圆桌： 郭源、刘亚奇、刘亚东、白承麟、韦松伟

▲ 2025. 3. 14

目录

CONTENTS

- 从大模型到智能体
- 大模型智能体的类型
- 智能体设计开发
- 澜舟智能体
- 圆桌讨论

01

从大模型到智能体

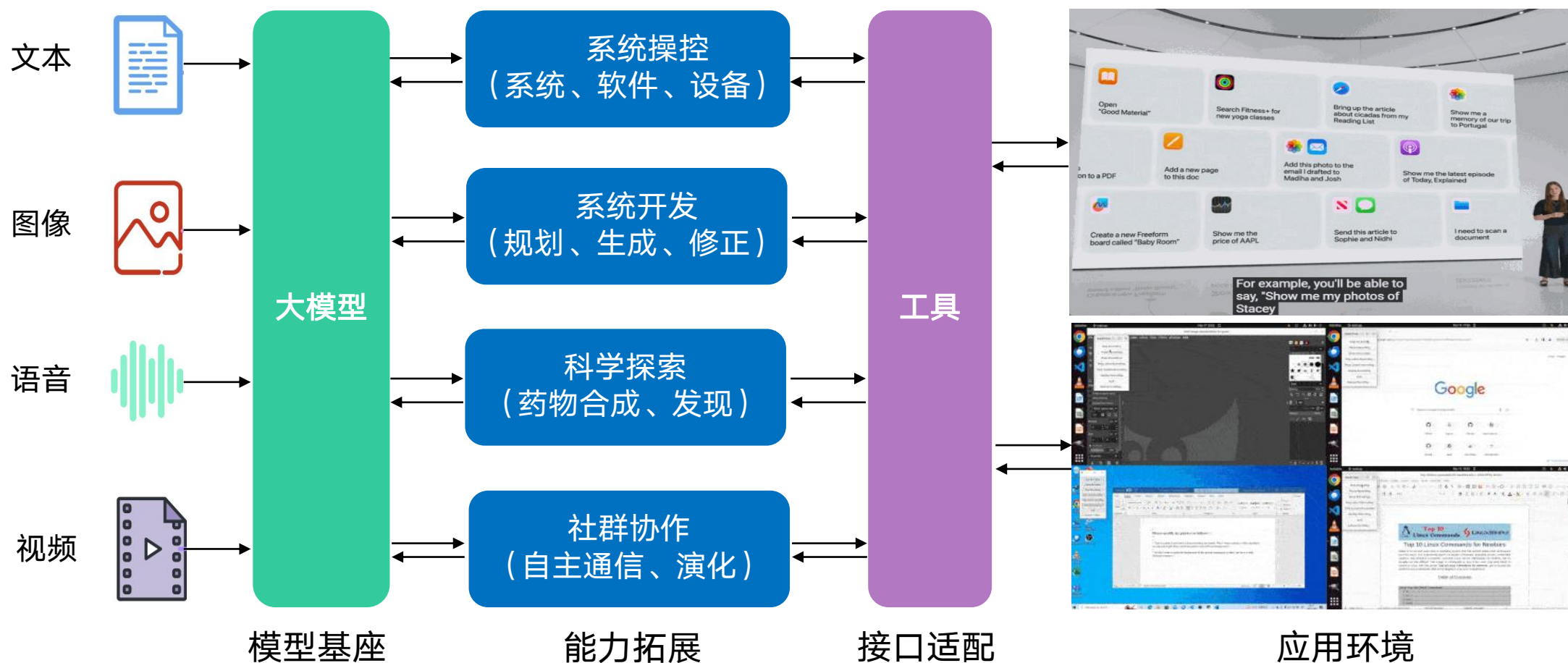




从内容智能到行为智能

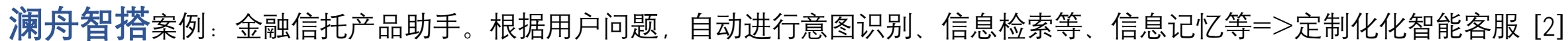
大模型在内容理解、推理和创作方向取得了显著的进展，但离**物理世界**存在着鸿沟

从**内容智能到行为智能**，构建面向开放环境的**大模型智能体**，建立迈向通用人工智能的关键纽带





「大模型智能体 ~ 大模型（大脑）+ 工具（手）+ 环境（世界）+ 记忆（经验）」

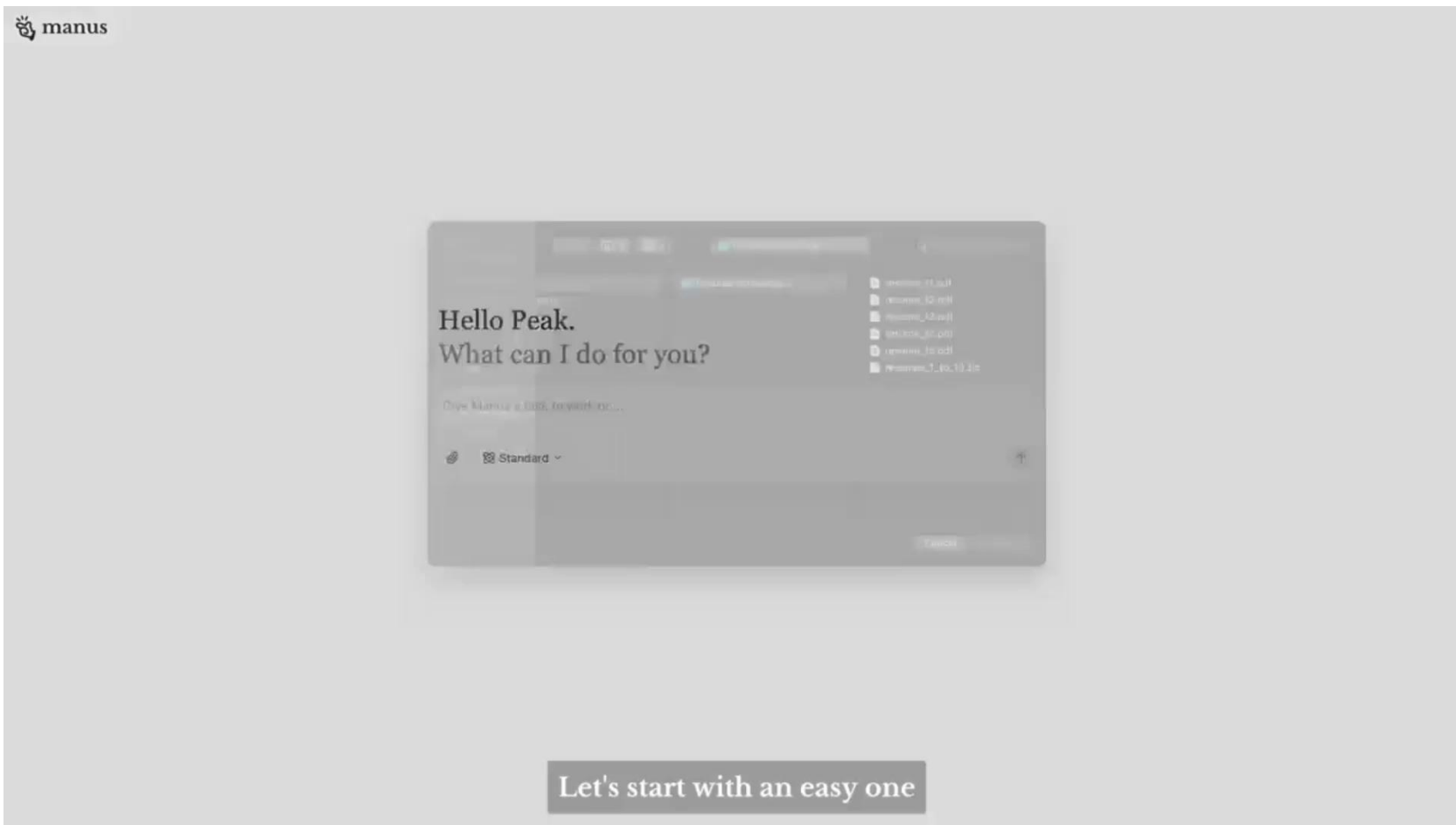


[1] What are AI Agents? IBM <https://www.ibm.com/think/topics/ai-agents> [2] <https://zhida.langboat.com/>



何为智能体？

「规划 + 感知 + 推理 + 决策 + 执行 + 学习」 大模型生成内容 => 智能体交付结果



“Service - as - a - Software”

Manus案例：用户上传简历压缩包，智能体规划简历筛选的流程，利用命令行工具解压压缩包，浏览各份简历，综合比对分析，给出详细对比报告。交付结果时根据用户指令调整展示格式，并记忆用户偏好。



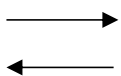
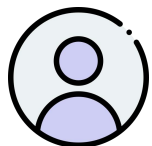
何为智能体？

人工智能（AI）智能体是指能够通过**设计其工作流程和利用可用工具**，代表用户或其他系统**自主执行任务**的系统或程序。AI 智能体可以包含自然语言处理之外的广泛功能，包括决策、解决问题、与外部环境交互和执行操作。

（多模态）大语言模型

用户

大模型



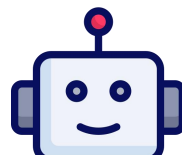
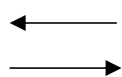
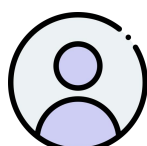
□ 通用能力的提供者

- 长序列处理
- 思维链推理
- 广泛意图理解

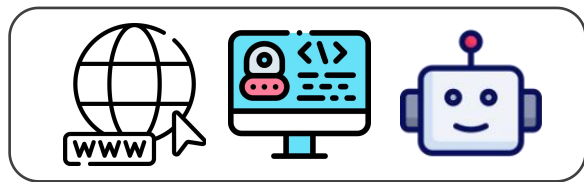
大模型智能体

用户

智能体



环境



□ 具备自主能力的智能体

- 环境交互：感知理解、规划决策、动作落实
- 持久运行：反思优化、迭代演化、安全防护

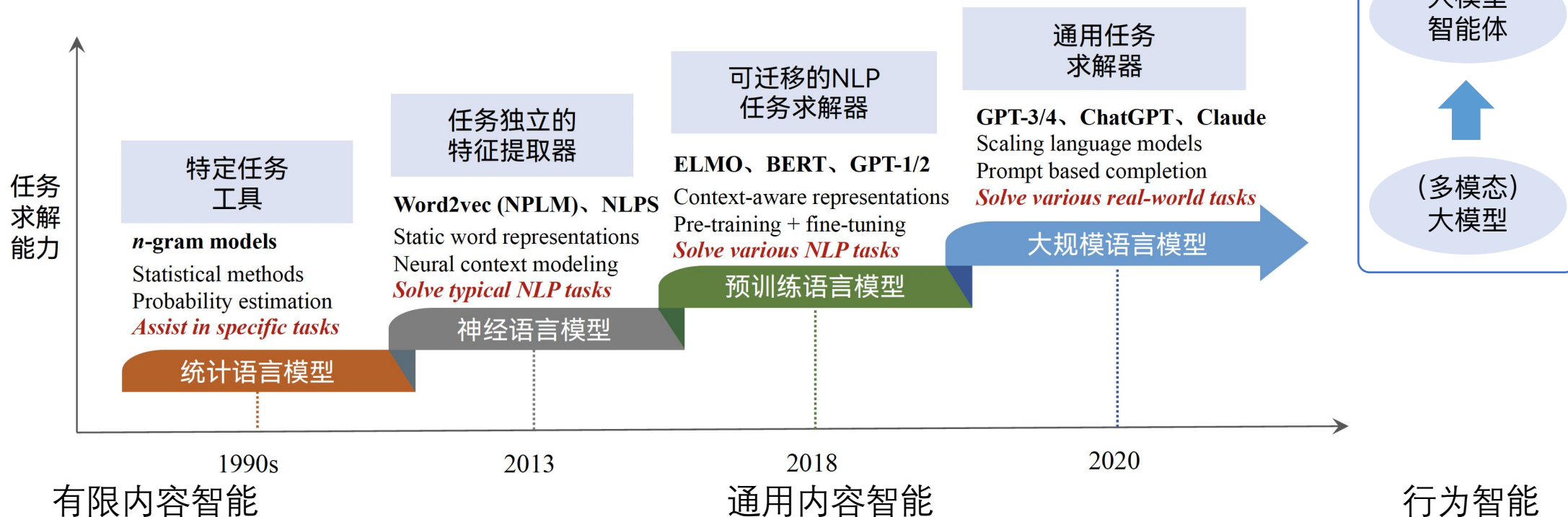
□ 知识更新：知识库、经验库

□ 集成工具：APP、API、函数



大模型智能体的发展背景

- 人工智能技术经历了小数据到大数据，小模型到大模型，专用到通用的发展历程，正逐步进入（后）大模型时代
- 在海量无标注数据上进行**大规模自监督预训练**，学习到大量的**语言知识与世界知识**
- **大模型智能体**的能力突破，成为新焦点，极大拓展大模型能力边界和应用场景





后LLM时代的智能体机遇



模型能力愈加成熟

大语言模型LLM→多模态大模型MLLM→推理大模型LRM

DeepSeek-R1: AIME 65%正确率; Codeforces优于89%选手



社会价值更受关注

单纯QA无法满足用户需求, killer app尚未出现

仅依赖基座模型无法做到垂直场景适配、复杂需求管理等



开发生态逐渐完善

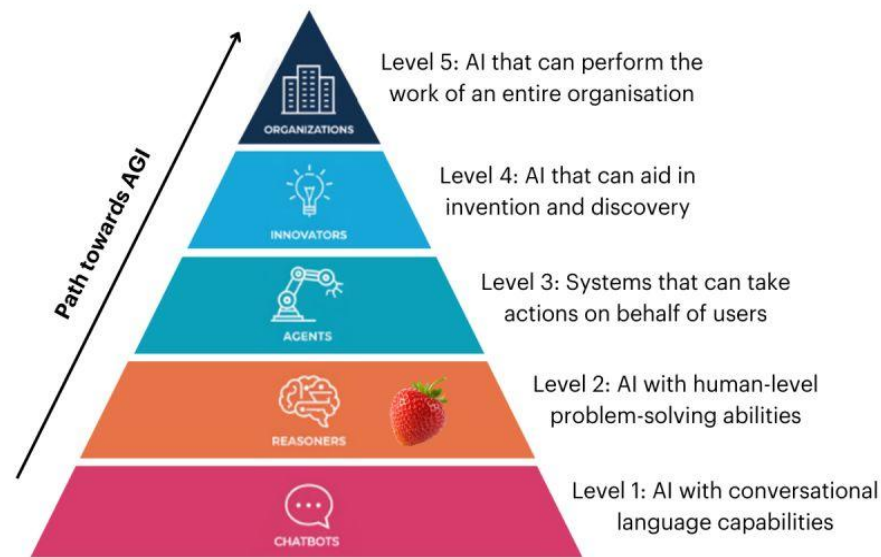
强大、开源的基座模型 (Qwen、DeepSeek、Llama)

完善的功能开发社区

(训练、部署推理、RAG、工具调用, 多智能体协作等框架)

The 5 Levels of AI

(OpenAI Classification System)



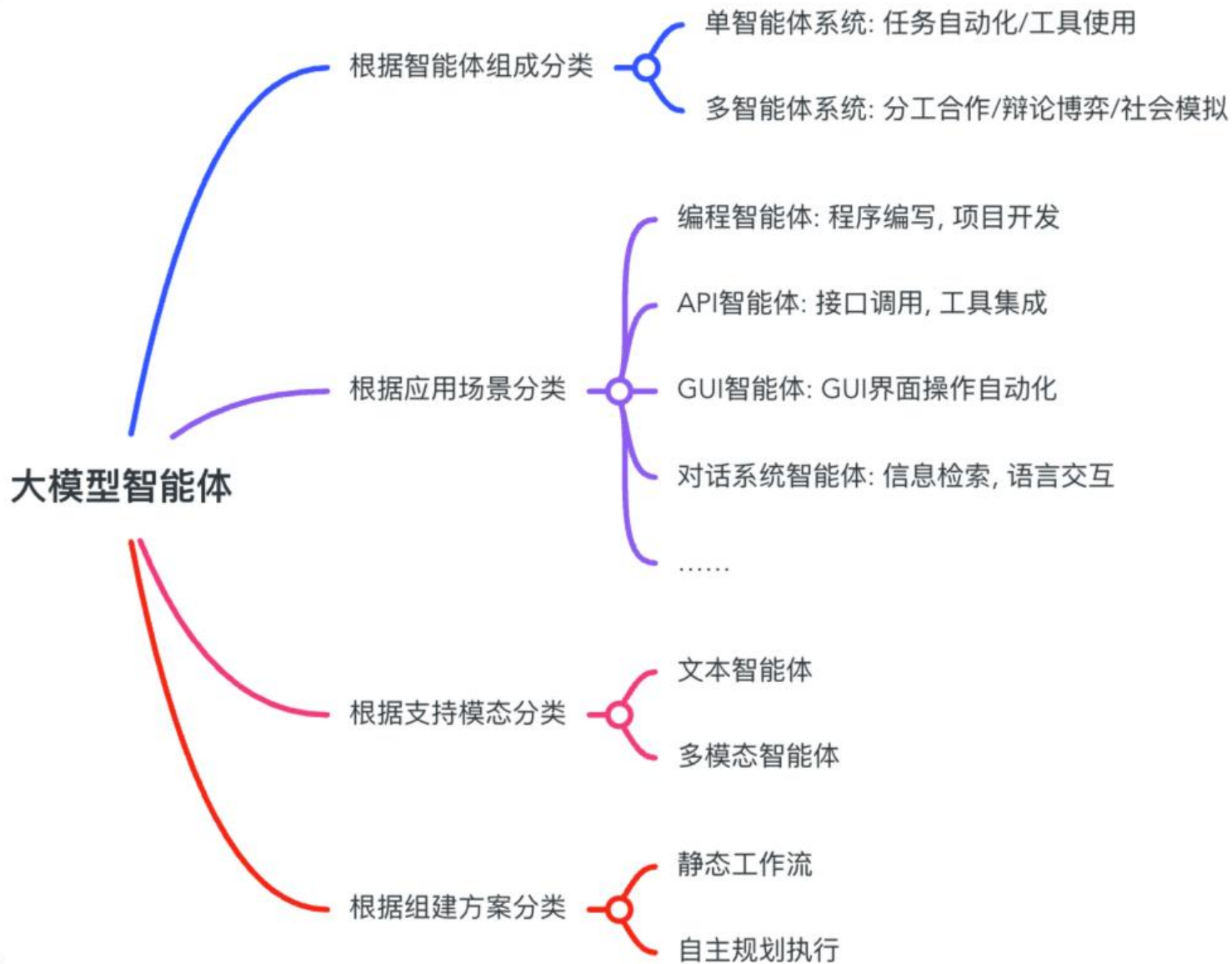
02

大模型智能体的类型





大模型智能体的类型



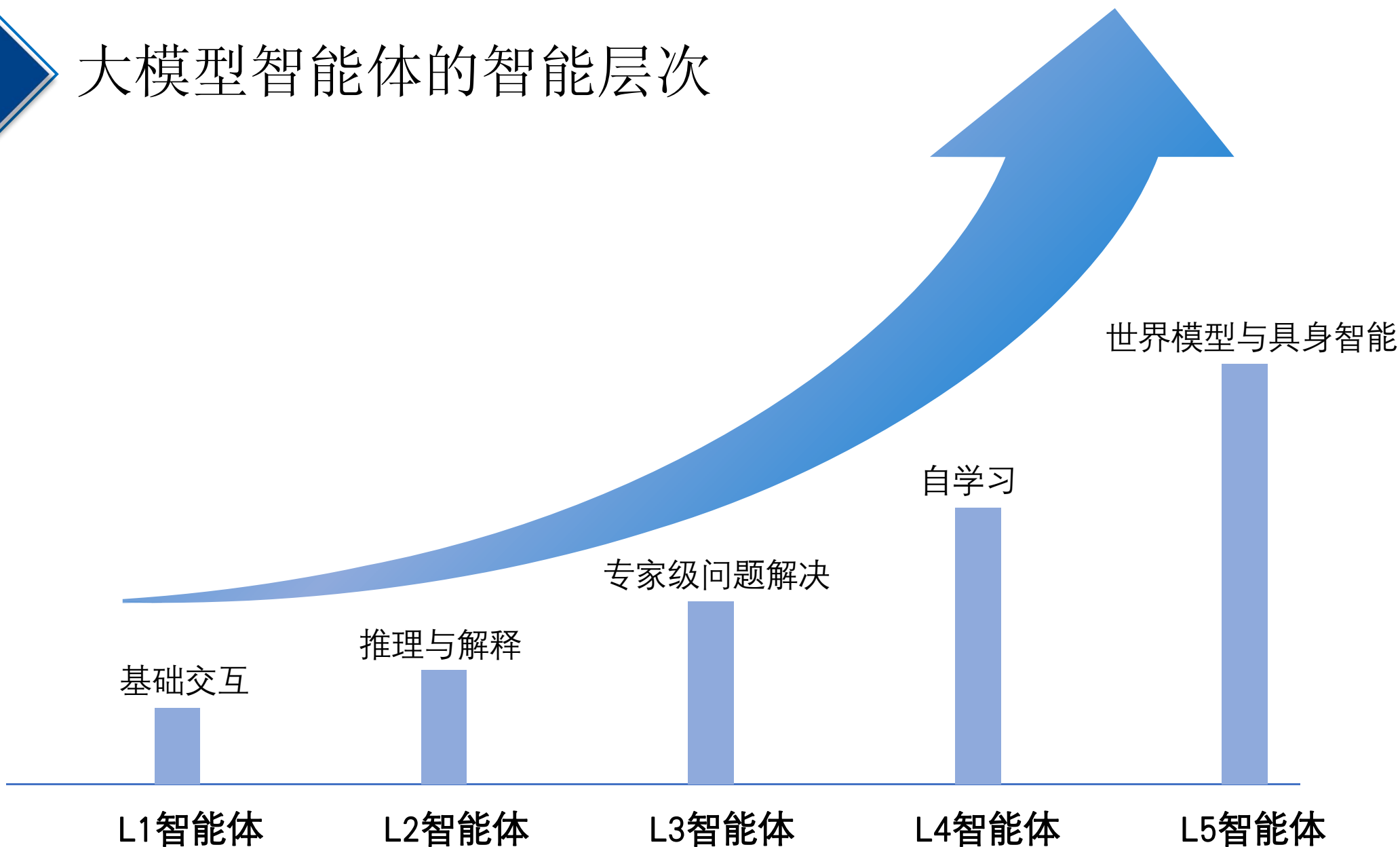
— 模块设计更加复杂

— 覆盖场景逐渐全面

— 能力区间不断延展



大模型智能体的智能层次





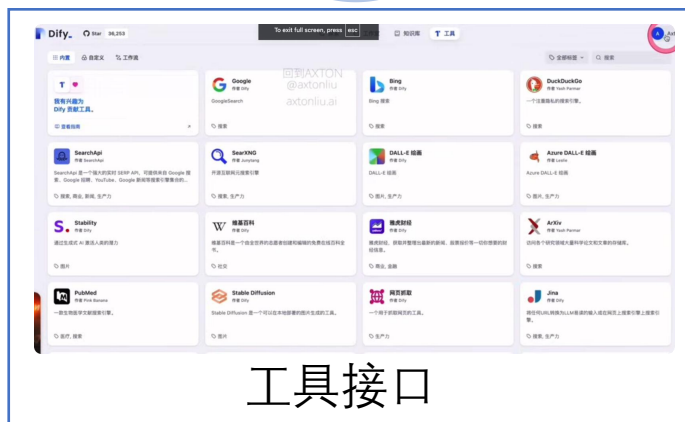
对话系统智能体

早期智能体应用建构的常见形态。在大模型的基础上增加外接知识库、工具接口、个性化设定等功能，从而应用大模型搭建复杂系统，提供定制化服务

案例分析：Dify框架（80k Github Stars的开源框架）



LLM APP



应用场景示例：本地部署的长文本一键翻译软件

文本分片
分片翻译
结果聚合
结果输出

优势：灵活的架构设计，突破了单纯获取response的上下文限制等问题

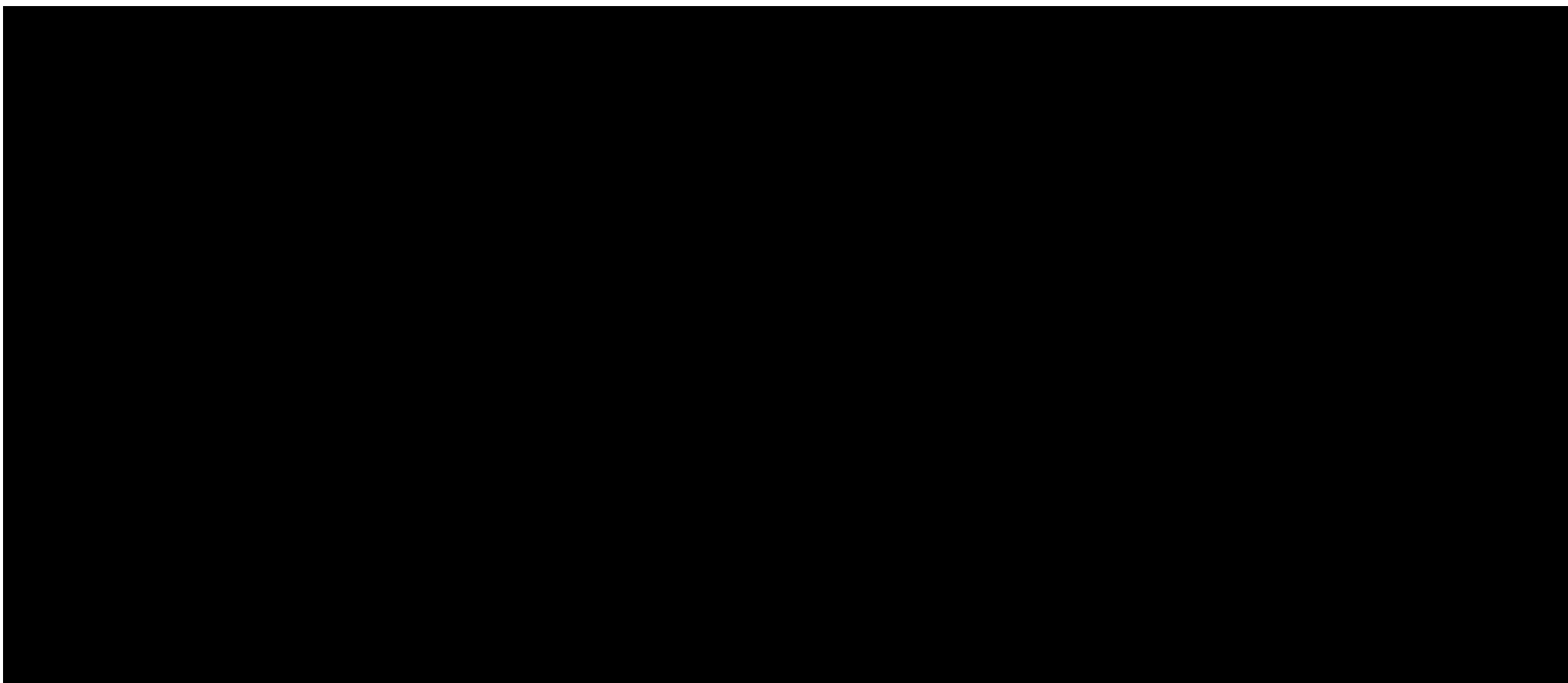


编程智能体

输入：编程问题/项目开发需求；一个代码项目仓库+一个issue（SWE-bench利用Github issue构建）

行为：开发计划；目录创建；程序编写，调试运行，bug修复，结果交付

“端到端的AI软件工程师”Devin 公司成立六个月估值20亿美元





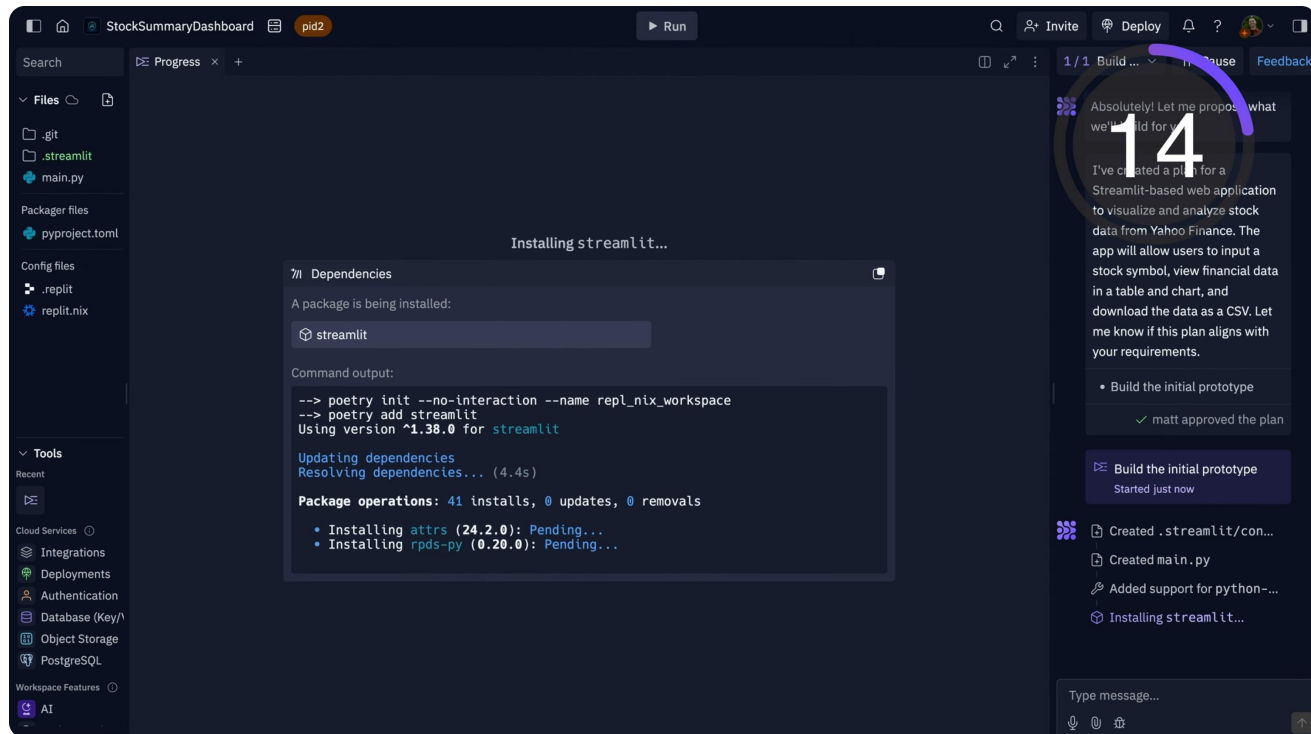
编程智能体

Replit: Idea to App, fast

针对“对编程一窍不通”的客户群体，不用写一行代码，就可编写项目

原理：基于Claude-3.5 API的多智能体协作

- 把复杂的开发流程拆分成子任务，分配给不同的智能体
- 将任务的执行过程透明化展示，使用户能够清晰洞察每一步的执行逻辑





编程智能体

复现思路：

1. 云端虚拟机控制，异步执行任务（类比：实习生） 2. 多智能体协作：任务规划+具体执行
规划智能体：拆分任务，制定计划 CodeAct：创建文件、编写代码，运行，debug

个人使用体验：

编程智能体已经发展较为成熟，使用Devin，可以全自动四小时内编写万行级别的手机应用并成功部署

Example

User:

① Find the sum of the reciprocals of the roots of $Sx^2-13x+4=0$.

Assistant:

② To find the sum of the reciprocals of the roots of the quadratic equation $Sx^2-13x+4=0$ [...]

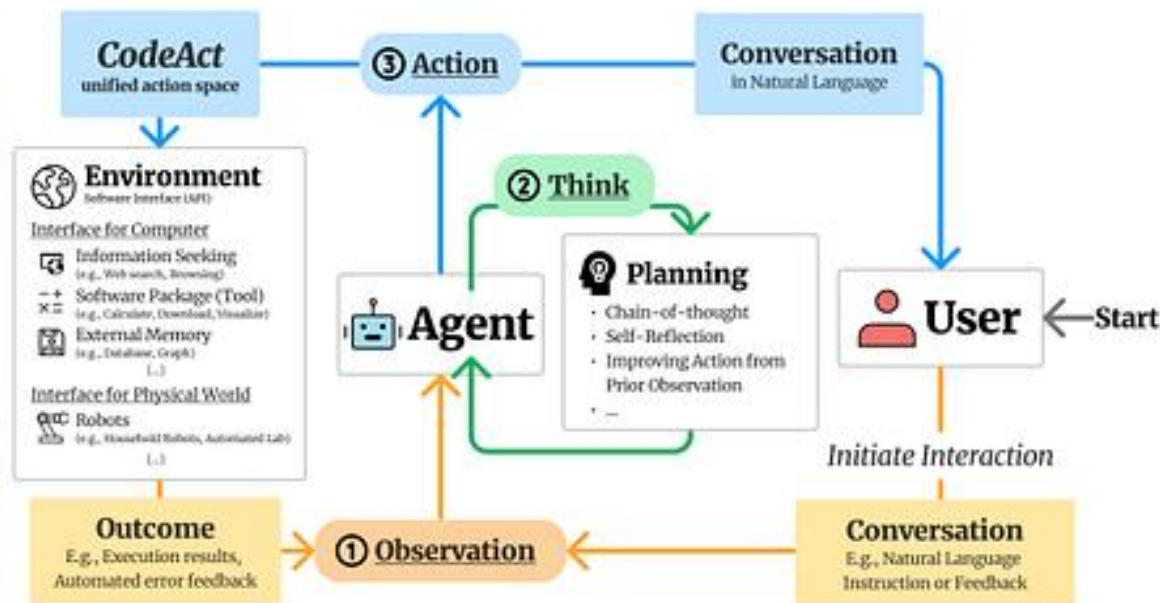
```
<execute>
import sympy
x = sympy.Symbol('x')
roots = sympy.solve(x**2 - 13*x + 4)
print(1/roots[0] + 1/roots[1])
</execute>
```

Env:

① $1/(3*\sqrt{17}/2 + 13/2) + 1/(13/2 - 3*\sqrt{17}/2)$

Assistant:

② The roots obtained are in symbolic form, which makes it difficult to evaluate the sum directly. Let's use the quadratic formula to explicitly find the numerical values of the roots first. [...]



半自动的类似产品

Cursor

WindSurf

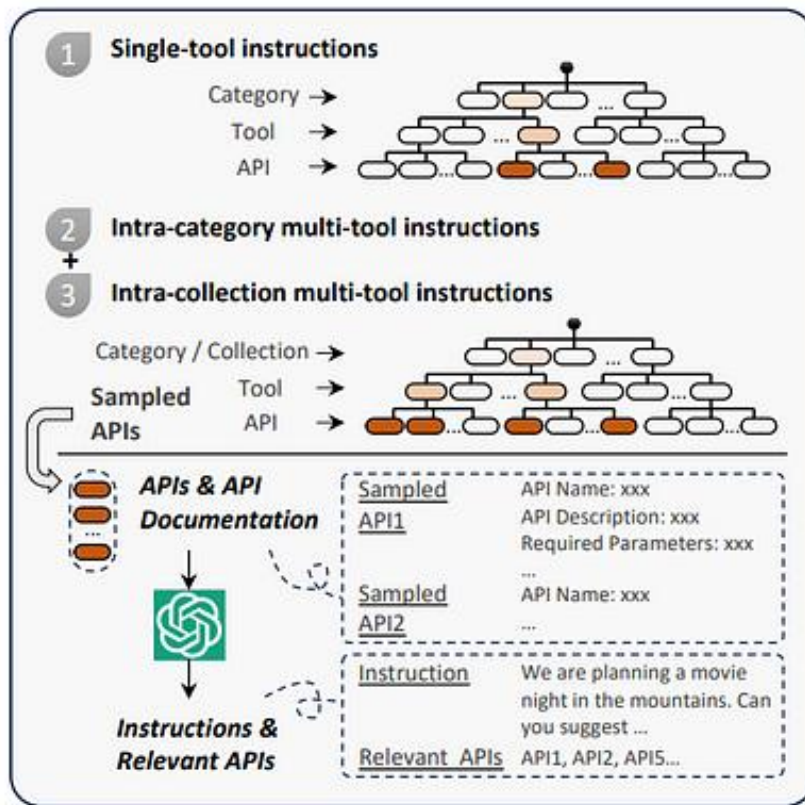
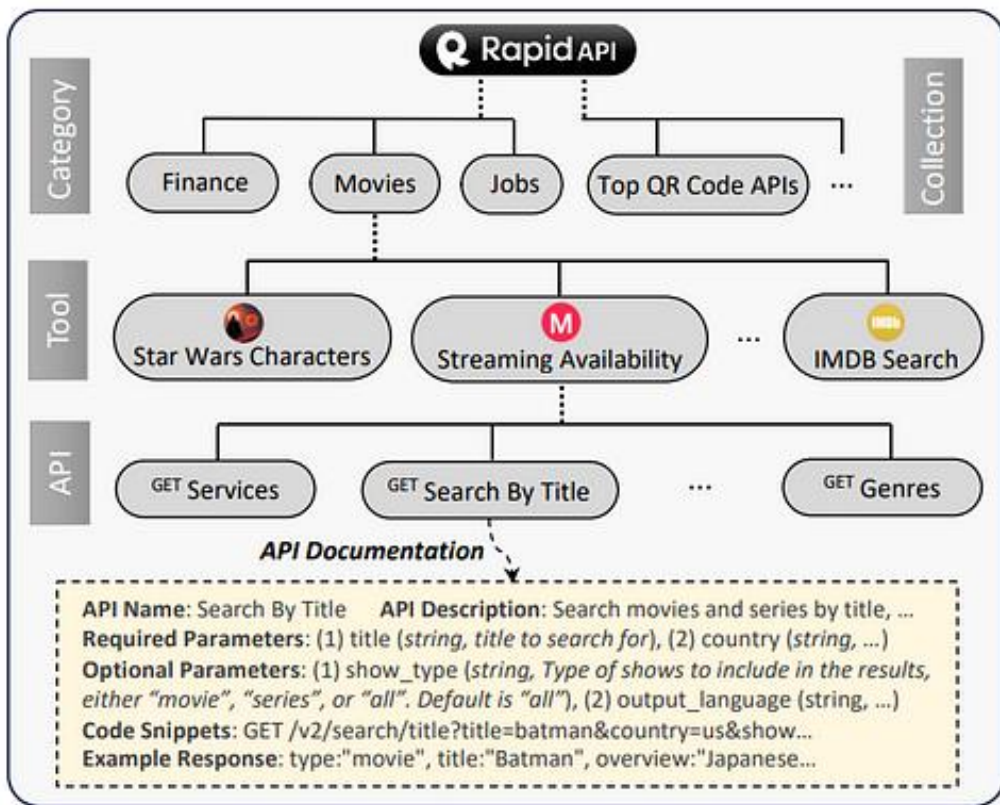
等



API智能体

利用大模型的工具使用（Tool-Use）能力，接入真实世界的API接口（工具）
在推理时，不仅利用自身的知识，还能领域调用外部api，扩展能力边界

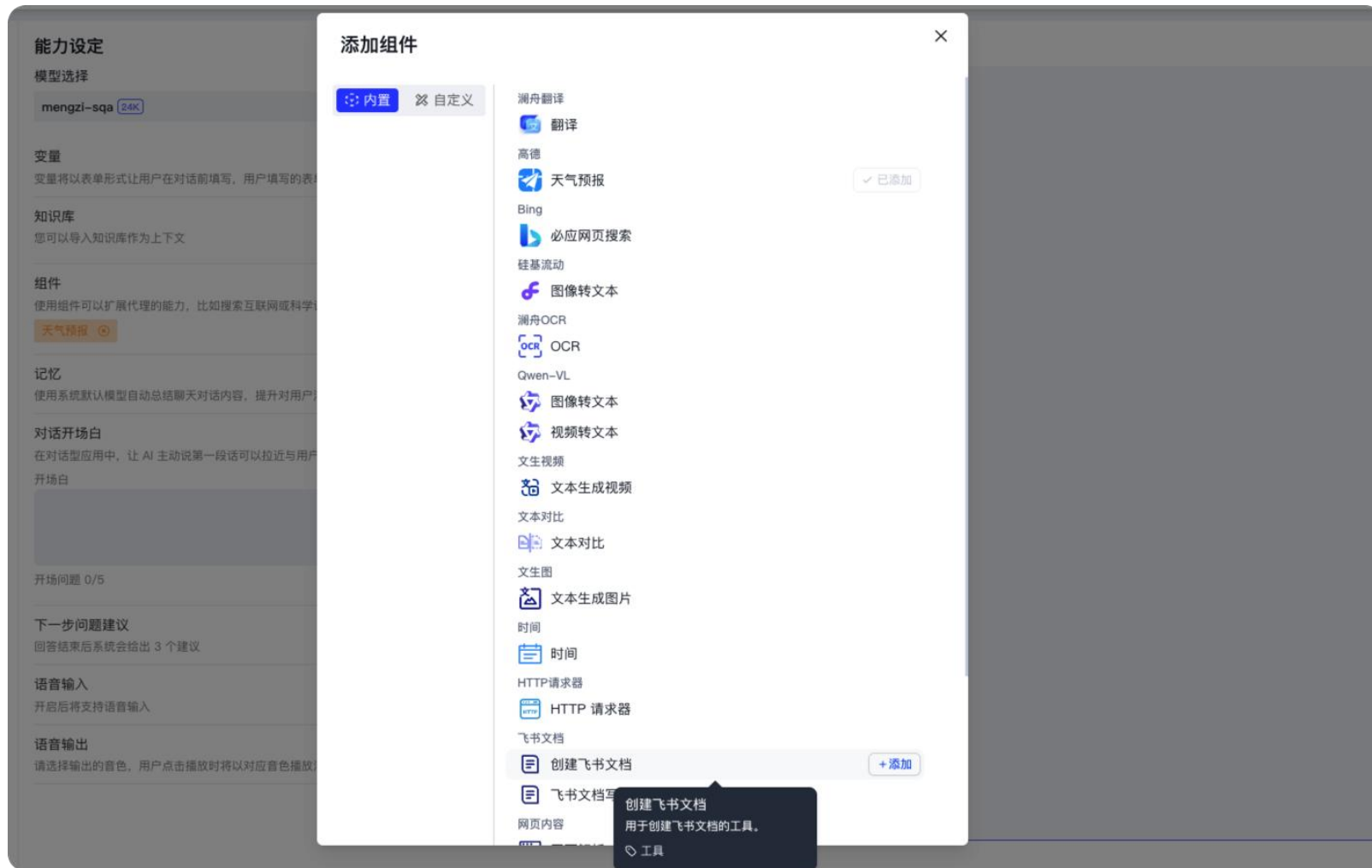
（如：股票分析问题—接入雅虎金融api；通用检索—谷歌api；领域具体业务—调用领域特定api检索；
车机自动操控—学习并使用车机的api）





API智能体

通过允许让智能体使用外界工具API，API智能体让智能体可以获得动态、实时、专有的信息源，提高决策能力；实现与外界的交互，拥有行动能力



澜舟智搭：
可以为自己创建的智能体灵活添加组件，使得智搭智能体决策过程中拥有调用特定工具的能力，更好适配业务场景
也可以添加自定义API

例如，搭建一个行程规划智能体，他在行程规划决策时可以调用高德天气预报API，获取出行日期天气信息，据此进一步安排出行计划。



GUI智能体

近期产品：OpenAI Operator智能体；Anthropic Computer Use；微软 OmniParser，UFO等

突破痛点场景，开辟新的交互范式

主要痛点：功能覆盖率低，尤其是第三方应用的接口尚未接入，严重影响用户体验

传统的API生态



播放周杰伦的《七里香》

```
{ "action": "play",  
  "song_id": "123" }
```



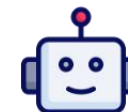
技术特点：API覆盖面有限、更新频繁难以维护、甚至缺乏开放API

能力特点：受限场景、特定的简单任务、固定操作流程

新型的GUI直接交互范式



播放周杰伦的《七里香》



技术特点：更具通用性、灵活性和拟人性，克服接口依赖，突破应用壁垒

能力特点：广泛覆盖面、复杂多样任务、适应动态变化



GUI智能体

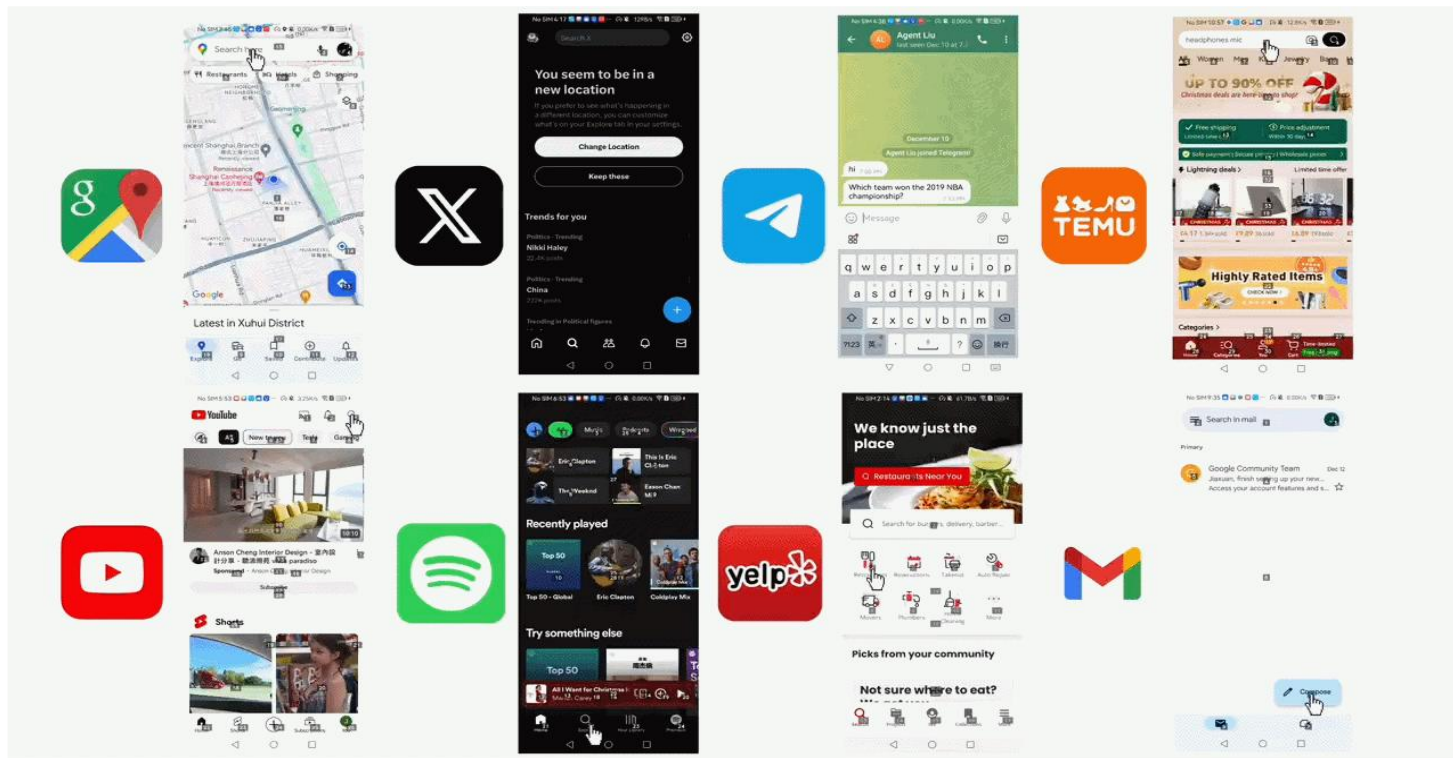
近期产品：OpenAI Operator智能体；Anthropic Computer Use；微软 OmniParser，UFO等

Operator：基于训练的CUA智能体，自主操控浏览器完成任务，如根据购物清单自动买好东西

Computer Use：教会Claude通用的计算机操作能力，实现拟人操作

UFO：Windows平台定制Agent，可以操控各种应用，实现PPT制作等任务；OmniParser：解析屏幕

开源框架browser-use：集成GPT接口，可根据用户指令操作浏览器



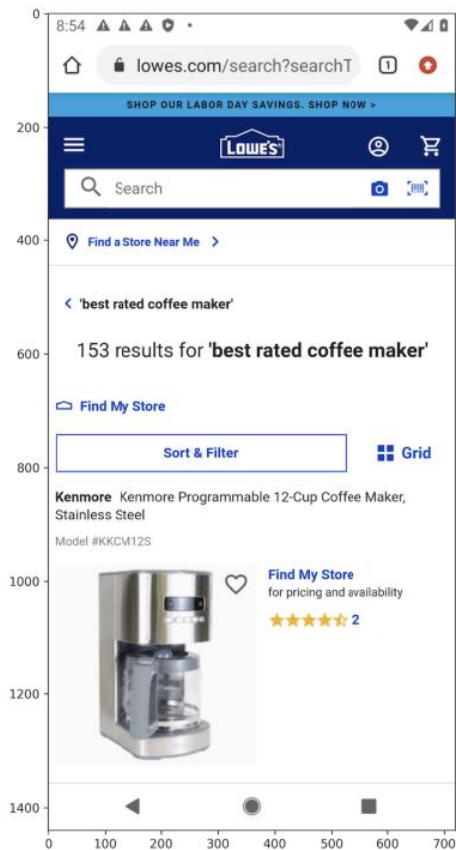
举例：

- 昨晚的外卖再来一份
- 打开随申码
- 去瑞幸小程序点杯美式送到办公室
- 订一张明天中午北京到上海的高铁票

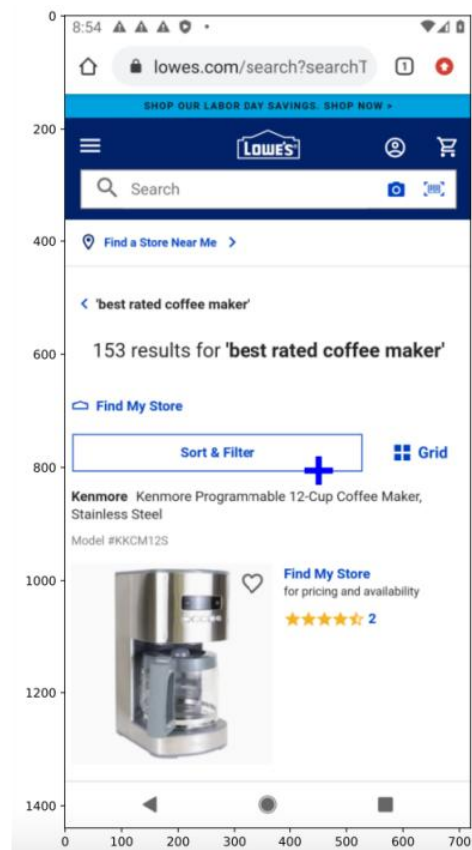
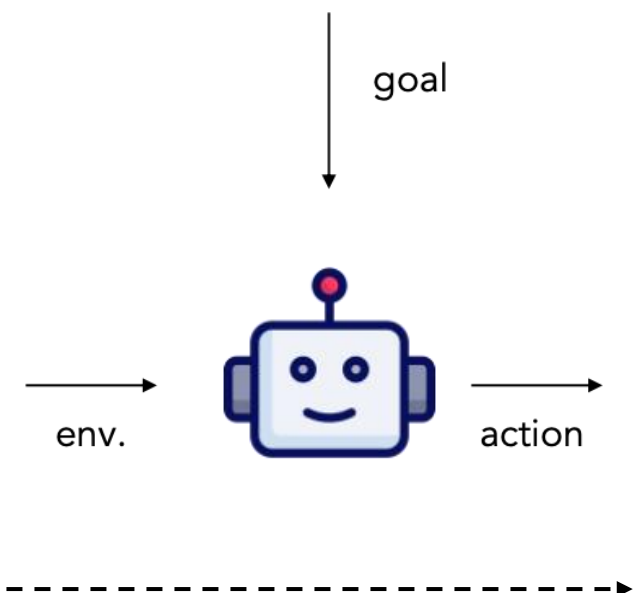


GUI智能体

Goal: Look up the best rated coffee maker



i -th screen



$(i+1)$ -th screen

输入：用户指令 + 当前屏幕 + 历史
输出：动作指令

Action Decision:

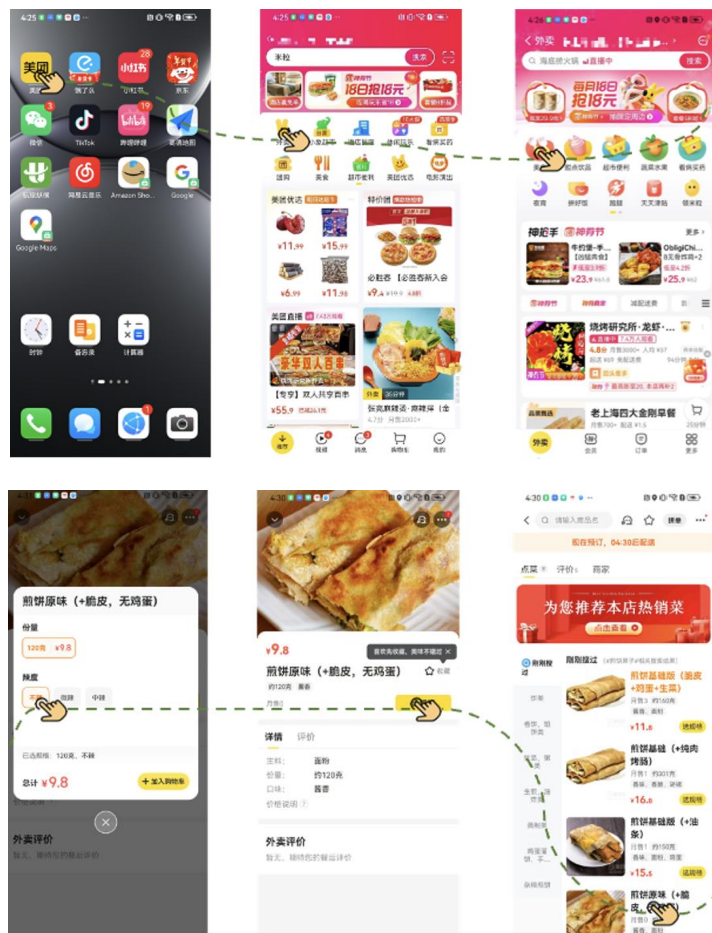
action_type: [DUAL_POINT],
touch_point: [0.5595, 0.6261],
lift_point: [0.5595, 0.6261],
typed_text: “ ”



GUI智能体

预训练（GUI定位能力）-> SFT（GUI操作常识）-> RL（能力外推）

任务指令：美团点份煎饼果子



感知理解

<页面概览>
<GUI布局>
<功能状态>
<图文细节>

规划决策

1. 打开“美团”APP
2. 搜索“紫燕百味鸡”
3. 选择最近的店铺
4. 下单送到默认地址

动作落实

动作类型：CLICK
坐标定位：<120, 351>
执行模式：自动

反思优化

异常识别

轨迹修正

迭代演化

自主探索

记忆演化

安全防护

对抗攻击

注入攻击

后门攻击

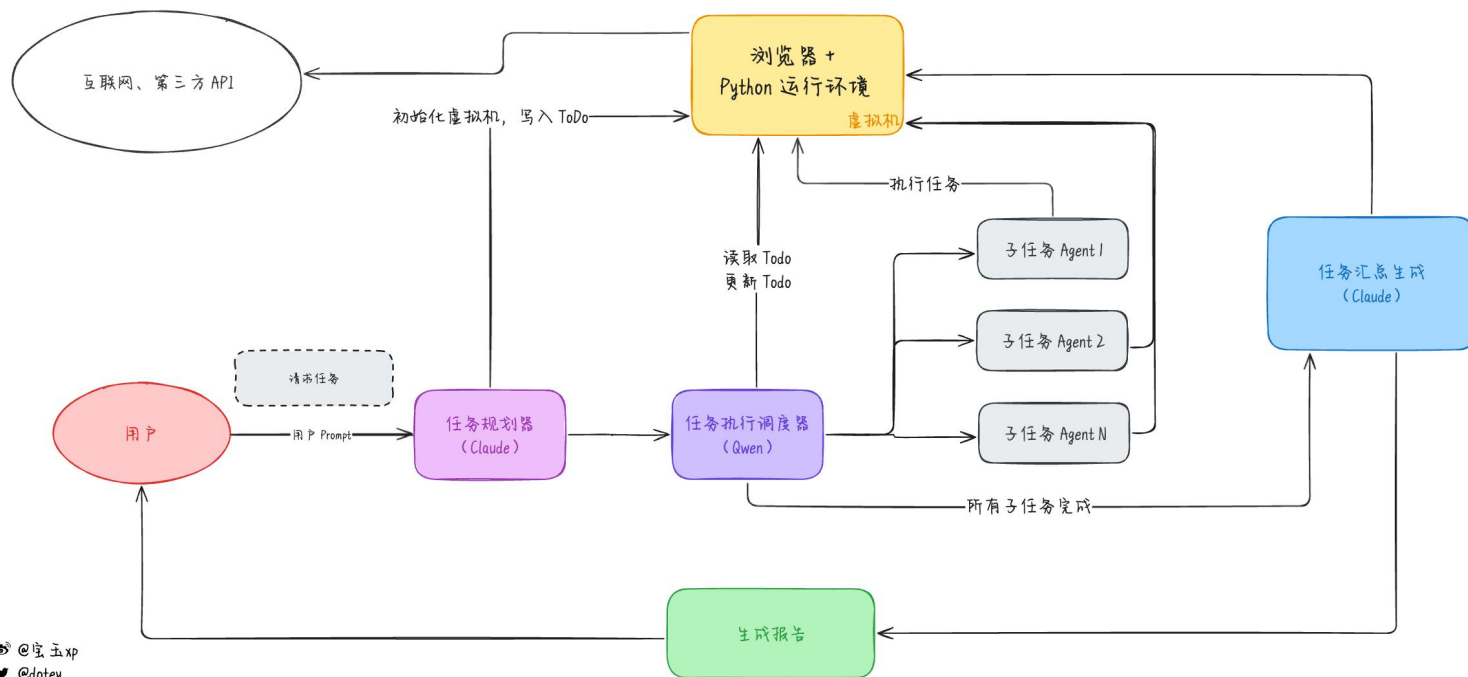
现阶段依然面临着感知、泛化、安全等方面的技术挑战，易过拟合、可靠性不足、易被攻击。通用场景可用性一般



解决复杂任务的多智能体调度系统

Manus; OpenManus; Owl等端到端智能体方案

Manus 的工作原理



@宝玉xp
@dotey

任务：你是个职业的二级分析师，擅长财务估值建模，建模精准、漂亮。请为英伟达做详细的excel建模估值。

任务拆解

Task progress 3/8

- ✓ 收集英伟达公司基本信息
- ✓ 获取英伟达历史财务数据
- Manus is working: 分析行业趋势和竞争格局 0:00 Thinking
- ① 构建财务预测模型
- ① 执行估值分析(DCF、相对估值等)
- ① 创建完整的Excel估值模型
- ① 编写分析报告和结论
- ① 整理并提交最终成果

灵活调度

- 检索财务数据：调用**雅虎API**等工具
- 分析产业格局：**Web**浏览相关报道
- 建立估值模型：编写**Python**脚本
- 编写分析报告：**通用**大模型

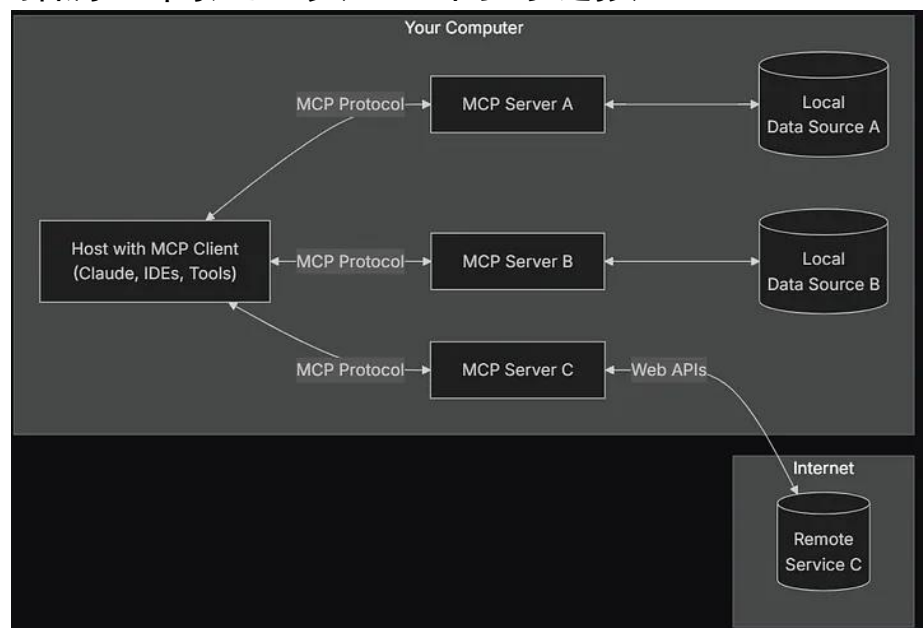
进度监督，信息汇总，结果交付



解决复杂任务的多智能体调度系统

MCP协议 (Anthropic)

MCP(Model Context Protocol, 模型上下文协议) 是一种新的开放标准协议, 用来在大模型和数据源之间建立安全双向的链接。



- (1) 统一的数据互联标准
- (2) 开放生态
- (3) 安全与权限管理

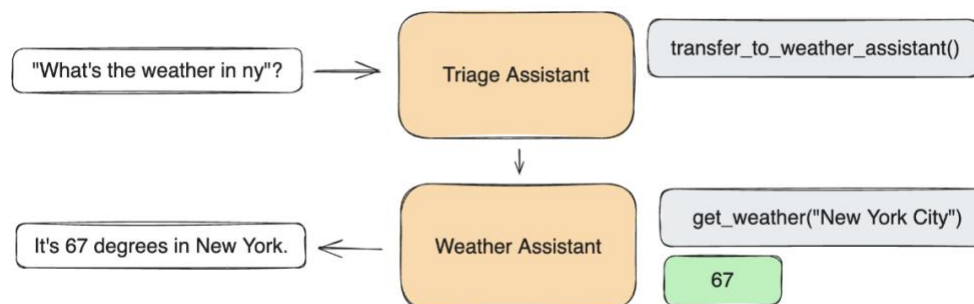
Swarm框架 (OpenAI)

专注于让Agent之间的协调和执行变得轻量化、易于控制且便于测试。

两类基本抽象:

Agent: Instructions + Tools

handoffs: 将对话交接给另一个代理



功效、轻量

更多的多智能体协作框架:

CAMEL, Magentic-One等

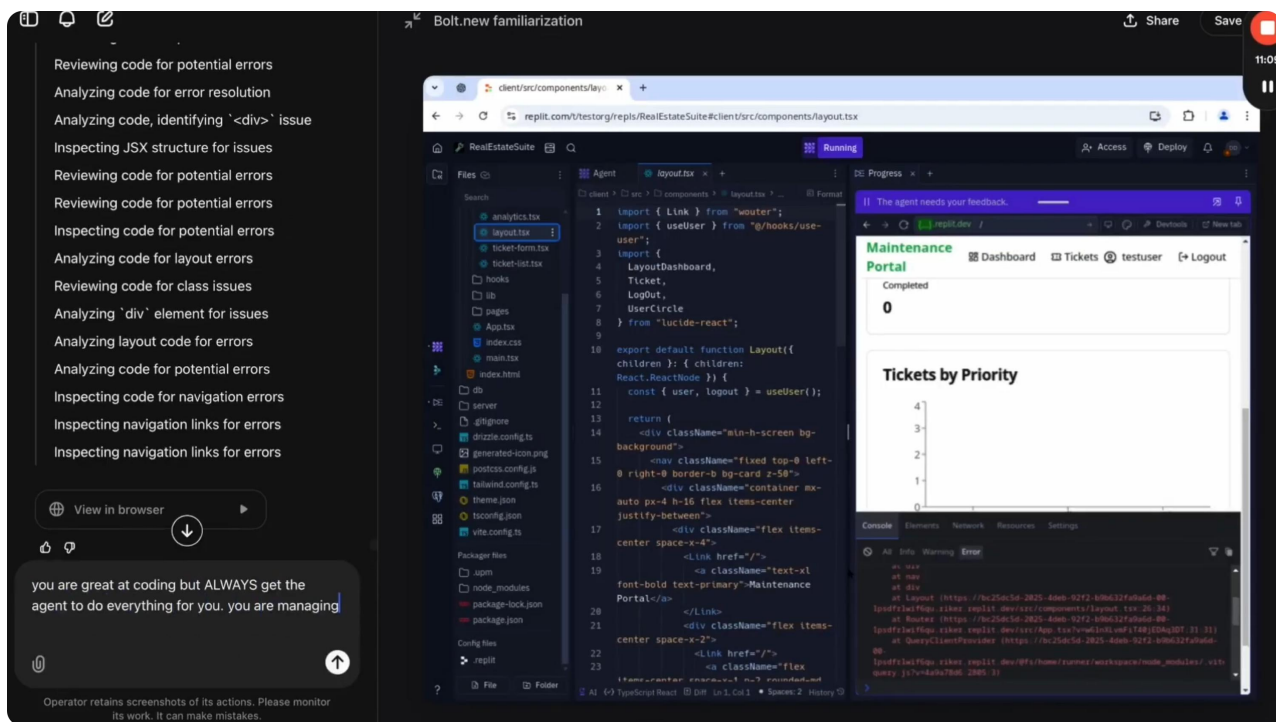


Takeaway: 大模型智能体发展现状

基座赋能：基座的蓬勃发展极大地拓展了智能体的想象空间（Replit）

覆盖场景：从简单的对于对话系统的增强，到面向开放环境，可以实现主动的环境交互、复杂任务规划与执行。相对而言，Coding Agent已经展现出不错的效果和使用价值(SWE-Bench: ~50%)，成为比较公认的生产力；整合API的工具使用智能体发展得比较早，在简单场景下也能比较好地使用；而基于GUI的智能体目前热度持续升高，但仍处于发展初期，有比较明显的任务执行表现和泛化能力不足的情况（OSWorld<40%）。

由单到多：基于多智能体协作的方案展现出前景。结合多元能力，完成复杂任务。这种多智能体协作不仅是完成任务中不同的“角色扮演”，也可以是不同职能的智能体的有机融合。



03

智能体设计开发





智能体搭建的两种理念

1. 端到端训练

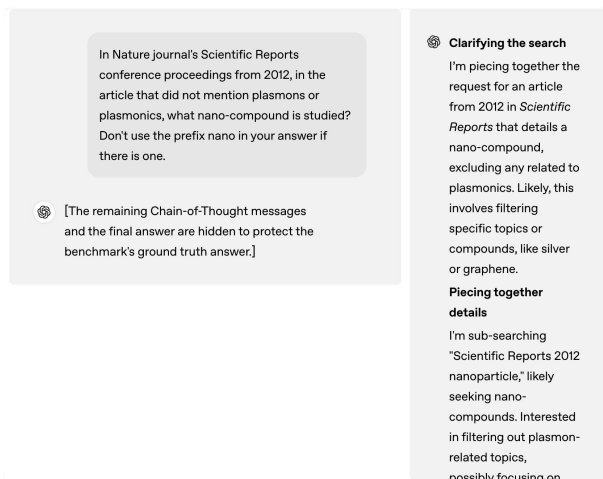
典型代表：OpenAI DeepResearch。通过端到端的强化学习训练，使得模型可以自发地组织出思维推理—搜索工具调用交织的分析思维链，原生地得出最终结果

优势：技术壁垒（RL）、强大的原生模型 缺点：训练困难；灵活性拓展性较困难

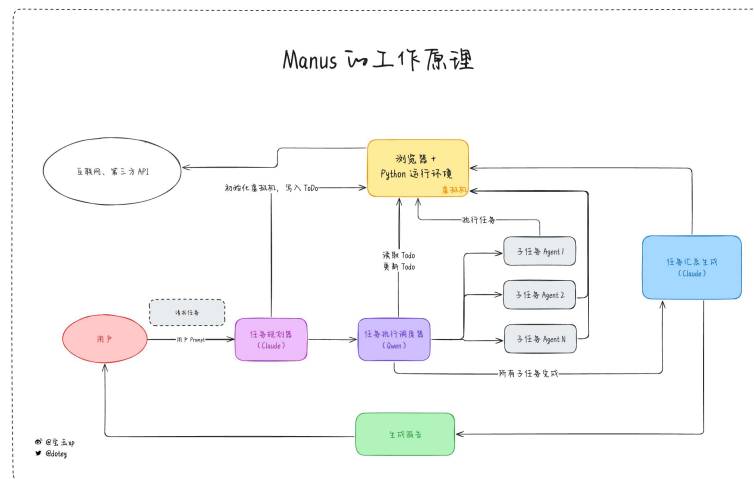
2. Wrapper和整合

典型代表：Manus。基本使用别人已经实现成熟的框架、模型，通过精巧的工程进行灵活的功能整合和集成效果；包括browser-use，也是对GPT层层封装，实现任务场景适配

优势：工作量小，站在巨人的肩膀上 缺点：（容易复现，架构设计需要精细）



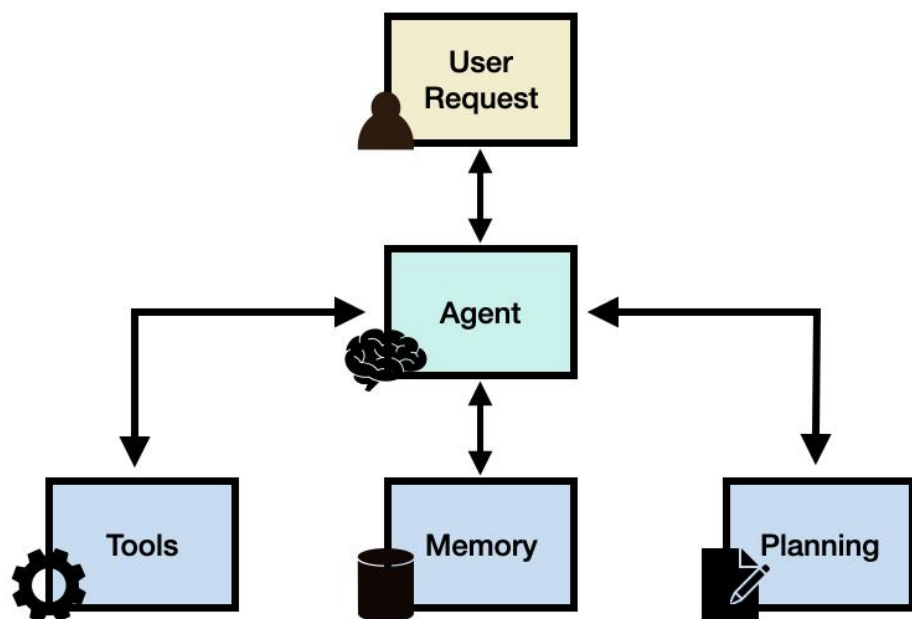
(a)



(b)



智能体生命周期



指令理解、意图识别和确认、恶意指令过滤

任务拆解，子目标规划，实时调整

工具配置（通用/个性化工具），何时调用

任务信息和用户信息记忆，智能体能力演化



智能体开发组件：OpenAI新发布的工具

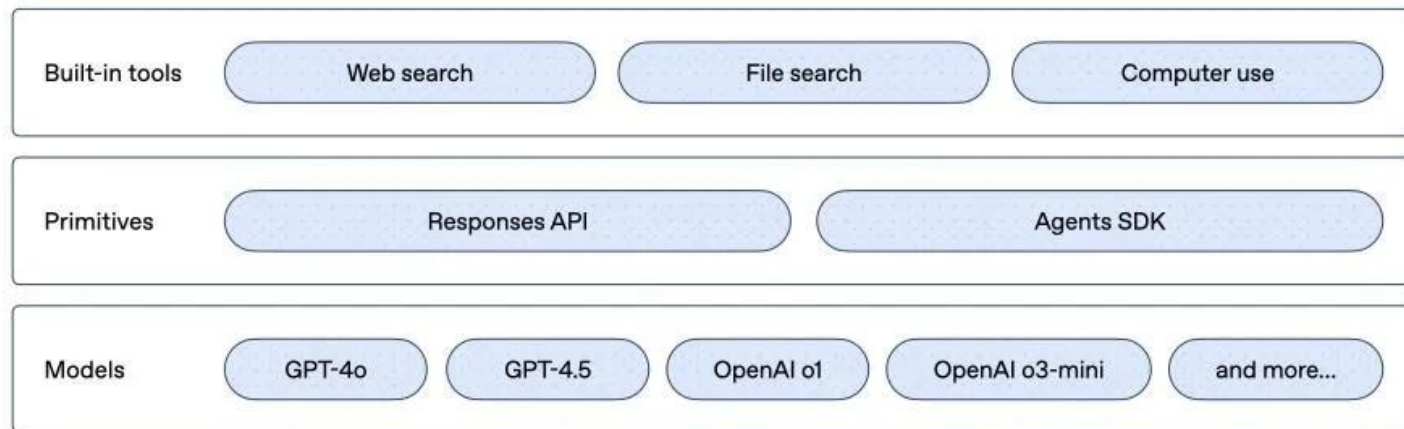
网络搜索工具：检索网络信息（基于GPT-4o-mini）

文件检索工具：上传文件，实现一键RAG问答。元数据过滤、直接搜索端点机制

Computer Use功能的直接接口，以及生产级的多智能体协作框架（SWARM的进一步）

（注：从这也能看出智能体构建的几个重要的功能和环节）

Agent building blocks





开源资源整理

Manus开源复现

<https://github.com/mannaandpoem/OpenManus>

<https://github.com/camel-ai/owl>

browser-use开源框架（操作浏览器执行特定任务）

<https://github.com/browser-use/browser-use>

<https://github.com/browserbase/open-operator>

开源DeepResearch（端到端地进行检索和思考分析，完成调研报告）

<https://github.com/mshumer/OpenDeepResearcher>

<https://github.com/assafelovic/gpt-researcher>

开源coding agent

<https://github.com/All-Hands-AI/OpenHands>

04

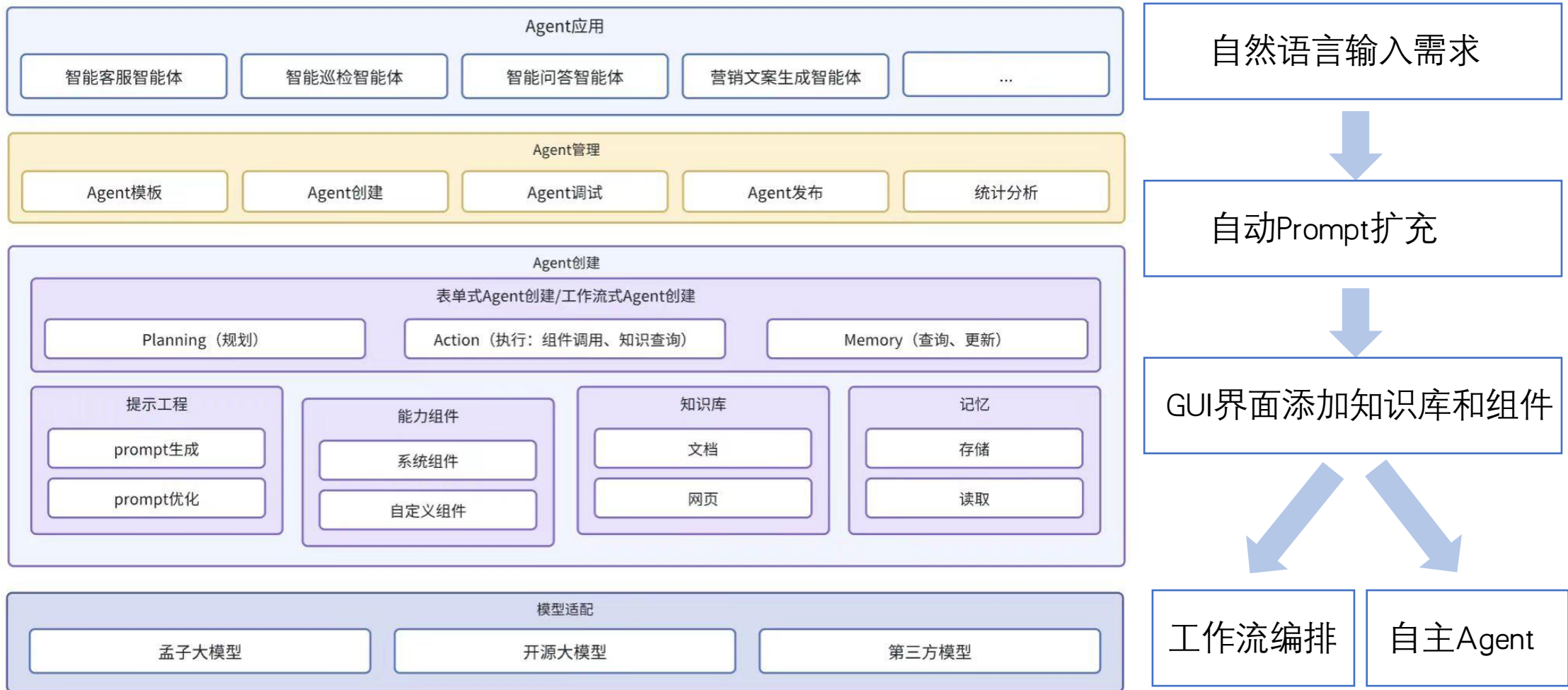
澜舟智能体





澜舟智搭

定位：帮助企业用户快速搭建贴合实际需求的智能体应用，优化业务流程，降低运营成本，加速业务升级
使用方式：两种方案（固定Workflow编排+agent），组件添加（常用基础接口+个性化数据库），GUI界面





澜舟智搭

- 提供示例应用（市场调研助手、智能客服等等）
- 可以结合基础组件库、知识库和自己的组件、知识库零门槛搭建领域智能体应用

+ 创建应用

示例应用 我的应用

智

智能搜问

智能文档问答应用，基于用户问题，检索知识库返回答案，实现澜舟智搭文档搜索问答效果

工作流

智

智能客服

智能客服应用，基于用户输入，进行意图识别判断，查询知识库返回或基于大模型自主能力回复，完成快速搭建智能客服

工作流

智

智能写作

智能文档写作应用，基于用户输入主题，查询知识库作为参考内容，完成文章大纲创作和文章内容创作。

工作流

销

销售话术建议

AI 销售专家随时在线，遇到沟通的问题随时找我。

工作流

金

金融信托产品助手

服务于多行业的客服问答场景，帮助企业快速建立产品智能客服体系，提升用户体验和客服人效。

工作流

跨

跨境邮件专家

撰写跨境出海营销邮件生成、回复和优化

工作流

诊

诊疗陪伴

诊疗陪练应用通过AI模拟患者，辅助医学生提升诊断技能，提供问诊评分，高效助力临床实践训练。

工作流

订

订餐客服陪练

客服陪练智能体模板，专注于帮助客服人员模拟和练习订餐流程，提升服务水平和应对能力。小美将扮演不同类型的顾客，通过对话场景模拟真实订餐过程，帮助客服人...

工作流

电

电商零售客服

擅长揣摩用户的情绪，用热情、专业的语气解答客户的各种疑

工作流

深

深度搜索

用户输入一个命题，从互联网上检索信息，判断用户的命题是否正确。

工作流

智

智能客服-质检智能体

对客服对话进行全方位的质量评估，并提供专业的分析报告和改进建议。

工作流

智

智能党建助手

党建汇聚力量，红色引领未来；凝心聚力，共筑辉煌。

工作流

旅

旅行规划助手

根据需求制定旅游规划方案

工作流

数

数据质检小助手

对用户评价进行分类/聚类打标，并且识别命中的关键词，进行分析并对分析结果置信度打分

工作流

市

市场调研助手

梳理调研主题：把抽象的调研主题，转化为可量化、可执行的调研主题；挖掘潜在需求：回到产品使用场景、品牌购买场景和心理转折场景去挖掘潜在需求；生成调研问...

工作流

市

市场报告查询助手

市场报告查询助手，通过搜索公开信息，快速生成行业市场报告，涵盖市场规模、产品结构等关键数据，助您初步掌握行业动态。

工作流

小

小红书营销文案助理

只要一个主题，帮你一键生成小红书营销文案

Agent

大

大厂运营面试官

互联网大厂的运营面试官，面试过上万个运营，储备了10w+ 高质量题库，全面考察候选人水平。

工作流

合

合同内容抽取

提取合同内的重要内容。

工作流

医

医美助理

医美智能助理，实现智能获取客户信息、产品介绍、术后陪伴、定期激活访问

工作流

产

产品卖点提炼

卖点提炼：抽取产品卖点； 大纲生成：基于提炼卖点生成PR稿大纲； 卖点对比：对比不同产品的卖点（当上传多个不同产品介绍时可使用）。

工作流

“

“金牌销售”车险续保

解决人工客服成本高、服务不及时、知识不专业等现实经营问题，提供智能化交互的销售服务体验，提升销售效率

工作流

S

SWOT分析

了解一个目标方向的优势、劣势、机会和威胁

工作流

P

PPT大纲生成

根据给定的主题，生成PPT大纲

工作流

5

5分钟深度行业研究

面向市场、咨询顾问、分析师和所有想要快速、深度了解某个行业的人员。提供要研究的行业，为你输出万字行业研究，包括行业概览（7维度）、4P分析（4维度）、行业...

工作流

发展展望

— 增加更多coding的功能支持，利用Python的广泛支持能力，外推服务范围边界（需要解决：文件传输、沙盒部署等）

— 复杂需求的规划能力，更多样的接口支持，用户信息的长期记忆等

04

圆桌讨论





圆桌讨论

1. 智能体时代，大部分技术都开源且同质化。（如强开源基座、全流程开发组件等）
技术逐渐平权的情况下，企业、产品的壁垒与竞争力如何构建？
技术？数据？客户？生态？
2. 目前已经出现了非常多比较有影响力的智能体框架、方案，但是大模型智能体在真实业务、生产中尚未得到大规模的落地和应用。这些框架距离最后落地的gap在哪里？
可能可以有效落地场景的大模型智能体方案、领域形态畅想？
3. DeepSeek-R1等高性能，低成本的推理模型的成功，为我们智能体的搭建带来了什么新的机遇？

Any Q & A

感谢倾听！
