

OpenAI Agents 常用代理模式

本目录包含了使用 OpenAI Agents Python SDK 实现的多种代理模式示例。这些模式展示了如何构建和组织 AI 代理系统来解决各种复杂问题。

目录

- [1. 确定性流程 \(Deterministic Flow\)](#)
- [2. 交接与路由 \(Handoffs and Routing\)](#)
- [3. 代理作为工具 \(Agents as Tools\)](#)
- [4. LLM 作为评判者 \(LLM as a Judge\)](#)
- [5. 并行化 \(Parallelization\)](#)
- [6. 防护栏 \(Guardrails\)](#)

1. 确定性流程

确定性流程是一种将复杂任务分解为一系列较小步骤的模式。每个步骤由一个专门的代理执行，前一个代理的输出作为下一个代理的输入。

主要特点:

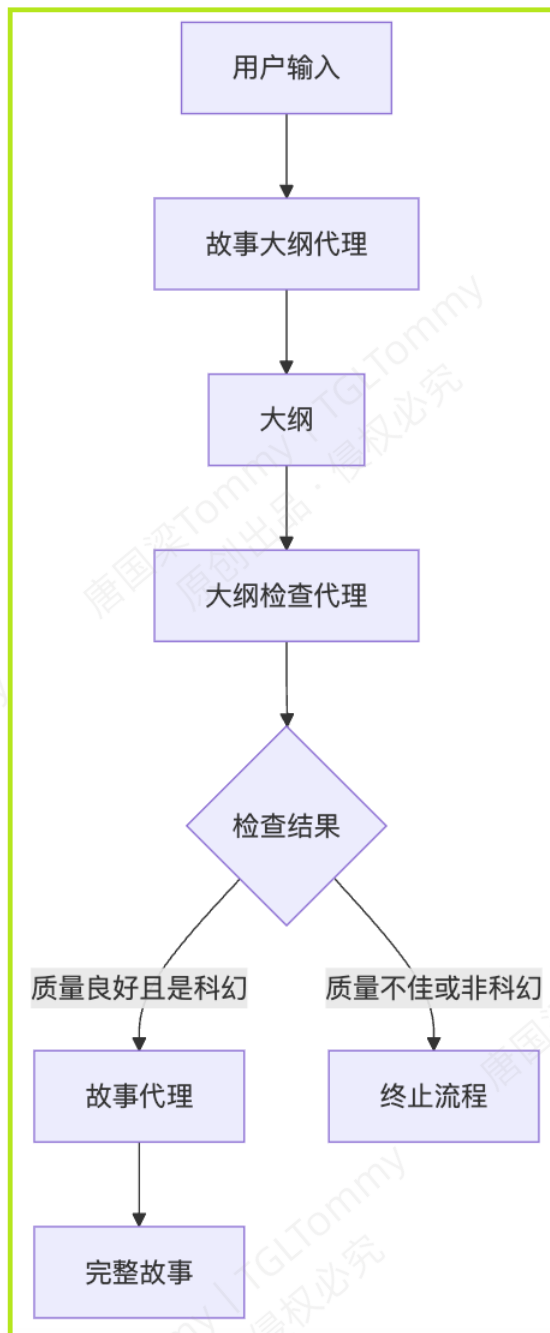
- 通过多个代理构建处理pipeline
- 前一个代理的输出作为下一个代理的输入
- 在步骤之间可以添加决策点
- 使用 Pydantic 定义结构化输出

示例实现: [deterministic_ollama.py](#) 展示了一个故事生成流程:

- 第一个Agent生成故事大纲
- 我们将大纲提供给第二个Agent
- 第二个Agent检查大纲是否质量良好，以及是否是科幻故事
- 如果大纲质量不佳或不是科幻故事，我们就此停止
- 如果大纲质量良好且是科幻故事，我们将大纲提供给第三个Agent
- 第三个Agent撰写故事
- 将最终的故事保存到本地文件

适用场景: 任何可以明确拆分为顺序步骤的任务，如内容创作、数据处理流程等。

流程图:



2. 交接与路由

在许多情况下, 你可能有专门处理特定任务的子代理。交接模式允许将任务路由到合适的专业代理。

主要特点:

- 前台代理接收用户请求并进行分类
- 根据任务需求将对话交接给专门的代理
- 接管代理可以看到之前的对话历史

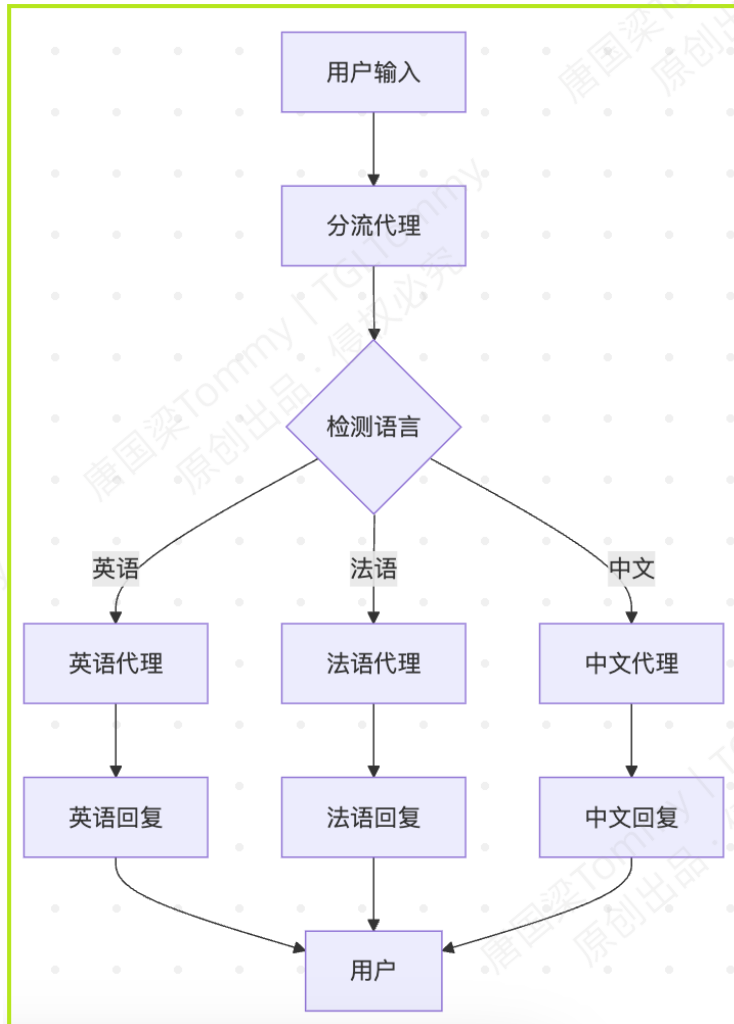
示例实现: [routing_ollama.py](#) 展示了一个语言路由示例:

- 分流代理模式: 使用一个主要的"分流代理" (triage_agent) 来判断用户使用的语言
- 语言专用代理: 根据判断结果将请求路由到专门的法语、中文或英语代理
- 基于Ollama的本地部署: 使用本地运行的Ollama服务作为底层大语言模型

- 实时流式响应: 响应内容实时流式返回, 而不是等待完整生成后一次性显示

适用场景: 客户支持系统、多语言服务、需要不同专业知识领域的应用等。

流程图:



3. 代理作为工具

与交接模式不同, 代理作为工具模式保持原代理对对话的控制, 仅调用其他代理作为工具来执行特定任务。

主要特点:

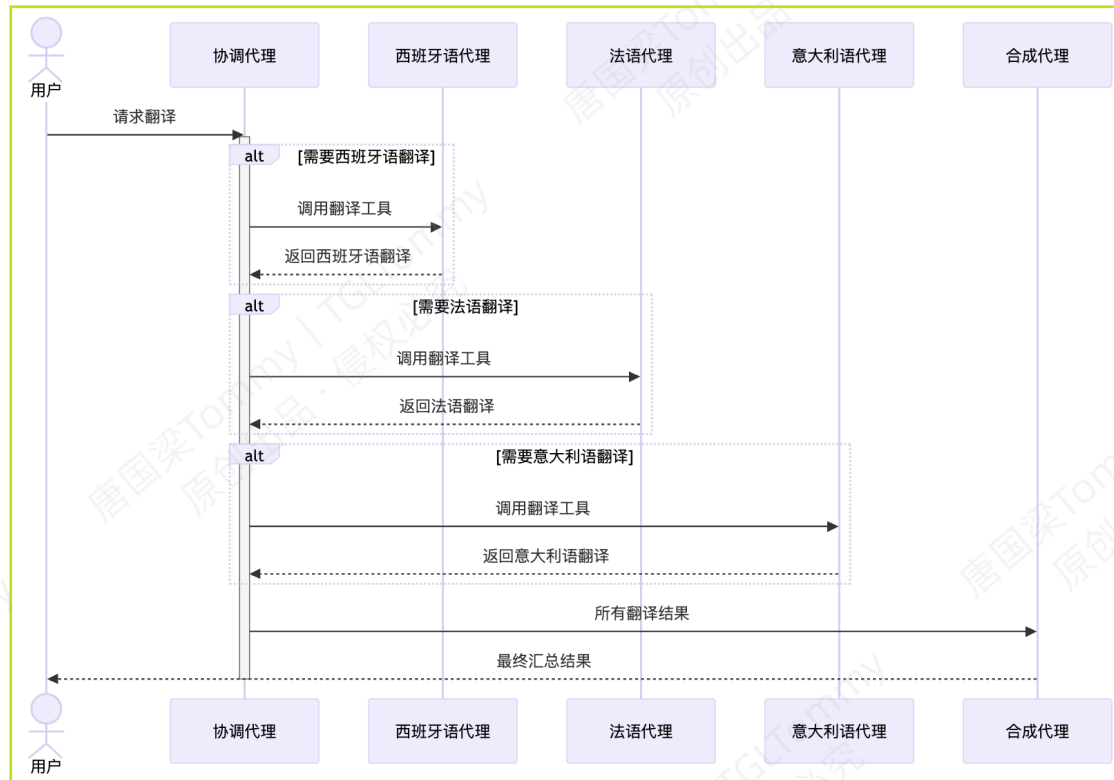
- 主代理保持对话控制权
- 工具代理独立运行并返回结果
- 可以并行调用多个工具代理
- 提供更灵活的工作流程设计
- 支持本地模型服务替代云API

示例实现: [agents_as_tool_ollama.py](#) 展示了一个多语言翻译系统:

- 主协调代理接收用户请求
- 根据需要调用不同的翻译工具 (西班牙语、法语、意大利语、中文或英语翻译代理)
- 合成代理汇总所有翻译结果
- 使用本地Ollama模型服务替代OpenAI API, 提供更灵活的部署选择

适用场景：需要组合多种功能的复杂任务，如多语言翻译、多模态分析等。也适用于需要本地部署、保护隐私或降低API成本的应用场景。

序列图：



4. LLM 作为评判者

LLM 常常可以通过反馈提高输出质量。这种模式使用一个模型生成内容，另一个模型评估内容并提供反馈。

主要特点：

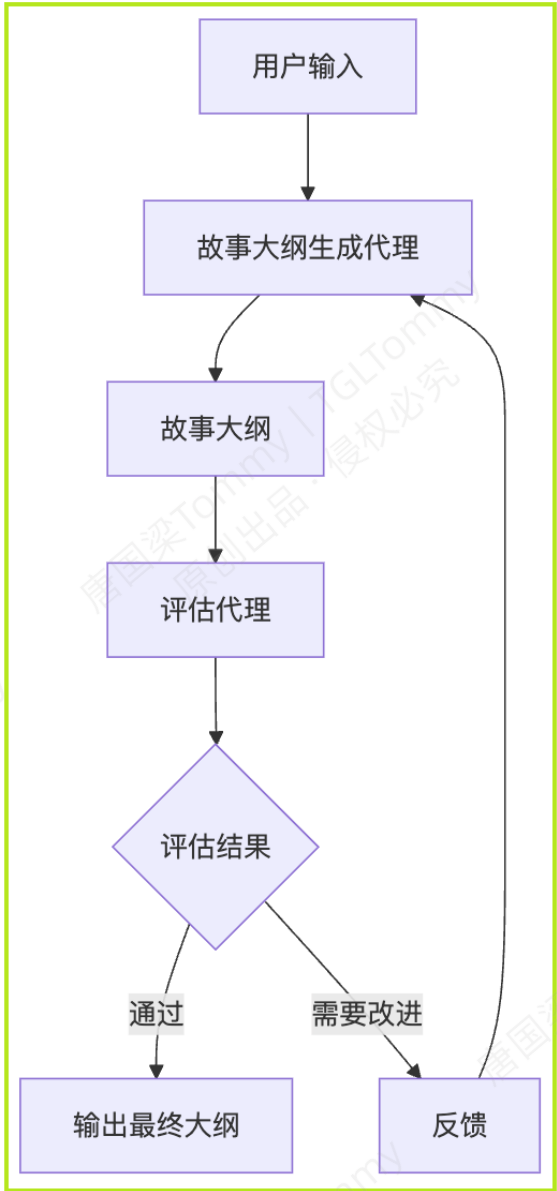
- 循环改进内容质量
- 使用结构化输出确保评估一致性
- 允许逐步优化生成的内容
- 可使用较小模型生成，大模型评估，优化成本

示例实现：[llm as a judge.py](#) 展示了一个故事大纲改进流程：

1. 第一个代理生成故事大纲
2. 评估代理检查大纲并提供反馈
3. 如果评估未通过，将反馈发回第一个代理
4. 循环直到评估代理满意为止

适用场景：需要高质量输出的任务，如内容创作、代码生成、报告撰写等。

流程图:



5. 并行化

并行运行多个代理是一种常见模式，可以优化延迟或生成多个候选结果。

主要特点:

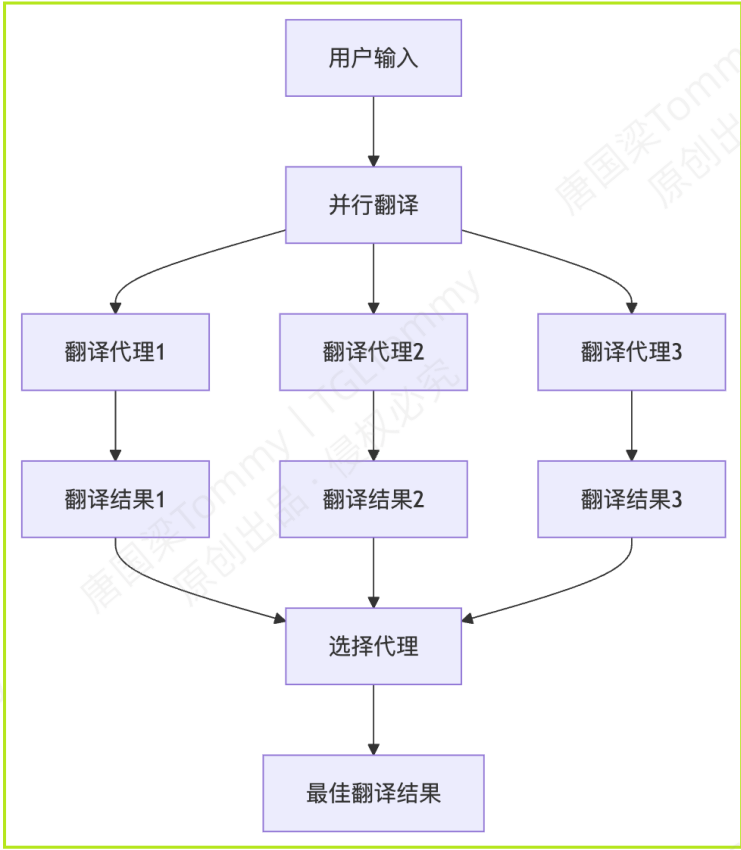
- 使用 `asyncio.gather` 并行执行多个代理
- 可以同时尝试多种方法解决问题
- 通过评选机制从多个结果中选出最佳答案
- 提高系统的响应速度和输出质量

示例实现: [parallelization.py](#) 展示了并行翻译系统:

- 同时运行多个相同的翻译代理
- 每个代理独立生成翻译结果
- 最后使用选择代理从多个翻译中挑选最佳结果

适用场景: 需要多种尝试的创意任务、对延迟敏感的应用、需要比较多个方案的决策系统等。

并行流程图:



6. 防护栏

防护栏是确保代理安全和合规的重要机制，可以应用于输入和输出。

输入防护栏

输入防护栏检查用户输入，确保其符合处理条件。

主要特点:

- 在代理处理前验证输入内容
- 使用"绊线"机制在问题出现时立即中断
- 可以快速拒绝不适合的请求
- 支持使用较快的模型进行初步筛查

示例实现: `input_guardrails.py` 展示了一个数学作业检测防护栏:

- 检查用户是否在要求代理完成数学作业
- 如果触发防护栏，立即拒绝请求并返回适当的消息

输出防护栏

输出防护栏检查代理生成的内容，确保其符合安全和政策要求。

主要特点:

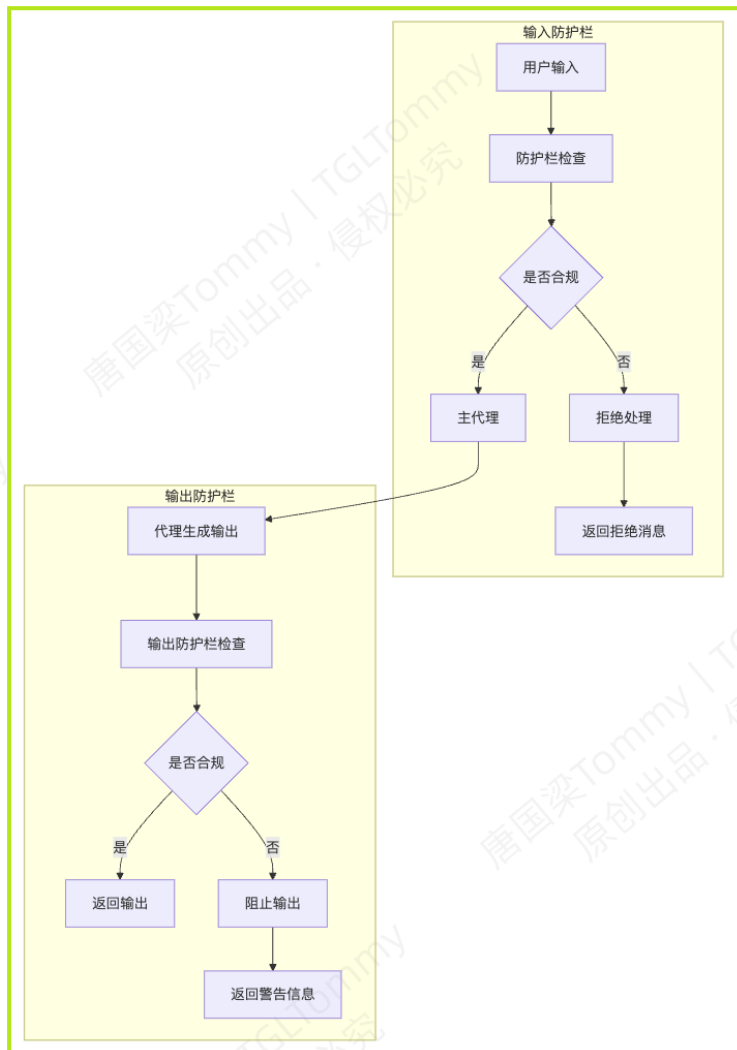
- 验证代理生成的输出内容
- 防止敏感信息泄露
- 确保回复符合内容政策

示例实现：[output_guardrails.py](#) 展示了一个敏感数据检测护栏：

- 检查输出是否包含电话号码等敏感信息
- 如果发现敏感信息，触发护栏并阻止输出

适用场景：需要内容审核的系统、处理敏感数据的应用、公共服务和教育平台等。

护栏流程图：



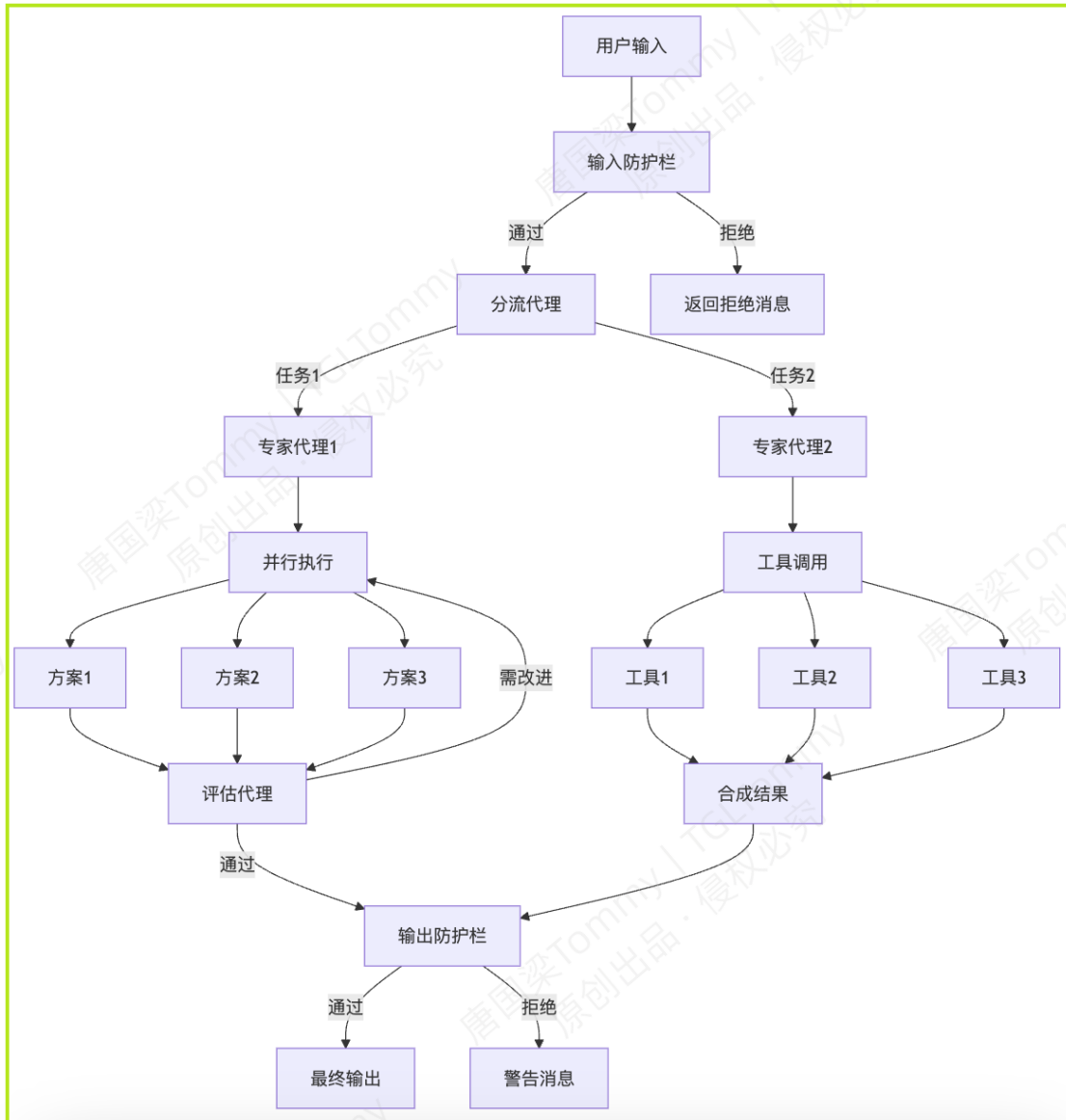
组合使用多种模式

实际应用中，这些模式通常会组合使用，创建更复杂、更强大的代理系统。例如：

- 使用确定性流程分解任务，同时每个步骤使用护栏确保安全
- 结合路由和工具模式，创建既有专业知识又能调用外部工具的系统
- 使用并行化和LLM评判者，生成多个方案并选择最佳结果

通过灵活组合这些基本模式，可以构建适应各种需求的复杂AI代理系统。

组合模式示例:



AI进阶学习

B站课堂: <https://space.bilibili.com/3546748815411393/pugv>

官方网站: <https://www.tgltommy.com/> (国内访问需科学上网)

如果你正在关注大模型 Agent、强化学习后训练、RLHF、DPO、GRPO、RLVR 等前沿方向, 欢迎关注我最上线的精品课程:

[《大模型 Agent 强化学习实战: 从后训练、奖励设计到工业级对齐系统》](#)

这门课程围绕当前大模型 Agent 强化学习的核心技术路线展开, 不只是介绍零散算法和论文概念, 而是希望帮助学习者建立一套完整的技术地图: 从基础原理、论文脉络、关键算法, 到开源项目源码解析与工业级系统实践, 逐步理解 Agent RL 为什么重要、如何演进, 以及在真实应用中如何落地。

课程配套专属飞书知识库《Agent RL 学习宝典》, 包含学习路线、论文资料、项目说明、源码阅读辅助资料和后续持续更新内容, 方便长期查阅和跟进前沿技术。

Agent RL 变化很快, 但越是变化快, 越需要一套系统的学习框架。如果你希望系统进入 Agent RL 方向, 而不是停留在碎片化了解阶段, 这门课会是一个很好的起点。

课程官网: <https://www.tgltommy.com/p/agent-rl> (国内访问需科学上网)

B站课堂: <https://www.bilibili.com/cheese/play/ss842375604>

更多课程详情, 扫描以下二维码咨询课程助理:

新课首发

大模型 Agent 强化学习实战:

从后训练、奖励设计到工业级对齐系统

唐国梁Tommy

TGL AI Tommy

官方微信课程助理

B站/公众号/YouTube: 唐国梁Tommy

**关注微信公众号
获取更多学习资源**



官方网站: TGLTommy.com(需科学上网)