何时（以及何时不）使用 AI 代理

预计时间：5 分钟

为什么评估是否使用代理很重要

AI 代理提供强大的功能，但并不总是最佳解决方案。在本阅读中，您将学习如何评估何时使用代理更为合理——以及何时更简单的工具更有效。您将探索现实世界的决策框架，并理解如何在创新与实用之间取得平衡。

学习目标

通过本阅读，您将能够：

- 确定代理在AI系统谱系中的位置。
- 应用四步决策框架来判断何时使用代理。
- 识别代理不适合的场景。
- 描述代理技术的当前局限性。
- 应用在部署代理时管理风险的指南。
- 理解有效代理架构的关键要素。

人工智能系统光谱

并非所有人工智能系统都在相同的复杂性水平上运行。以下表格描述了不同之处的光谱。

类型	描述	最佳使用案例
简单的人工智能特性	执行诸如分类或文本摘要等任务	快速、可重复的任务，具有明确的输出
协调的工作流程	预定义的多步骤逻辑，结合不同的人工智能特性	结构化流程，如文档审核或路由
自主代理	独立做出决策并适应新信息	复杂推理、探索或策略任务

接下来，了解可以用来评估何时使用代理的四个标准框架。

使用代理的四项标准框架

在构建或部署 AI 代理之前，请问自己以下问题：

1. 任务是模糊的还是可预测的?

当任务模糊时使用 *agents*：

- 决策路径不明确或无法提前规划
- 任务涉及探索、故障排除或创造性

当任务可预测时使用 *workflows*：

- 你可以定义所有规则和结果
- 该过程遵循明确、可重复的结构

2. 任务的价值是否值得成本?

由于探索开销，AI代理的运营成本更高。它们可能消耗 **10 到 100 倍** 的令牌，比工作流更多。让我们来看看一些场景：

场景	推荐
高投资回报率的战略规划	使用代理
基本客户支持任务	使用工作流代替

3. 代理是否满足最低能力要求?

在发布之前，测试代理的三到五项关键技能。

以下是一些示例：

- 研究代理必须识别、过滤和总结可信来源
- 编码代理必须编写、修复和验证代码片段
- 客户支持代理必须分类问题、解决常见查询，并适当地升级复杂案例
- 数据分析代理必须清理数据集、检测异常并总结关键趋势

如果代理未能通过这些测试，**缩减或重新设计代理**。

4. 如果代理出现错误，会发生什么？

评估这些问题的答案

- 你能迅速发现并纠正错误吗？如果可以，那么使用代理可能是合适的。
- 如果遗漏了什么，风险是什么？遗漏答案的后果会影响客户或组织的福祉或安全吗？
- 代理是否包含内置的纠正或验证工具？
- 在风险可控或可逆的情况下使用代理。

当前 AI 代理的挑战

即使是强大的代理也面临挑战。

挑战	重要性说明
推理不一致性	代理可能在一次任务中成功，但在类似任务中失败
不可预测的成本	资源使用可能会因复杂性而激增
工具集成问题	代理需要良好集成的工具和稳定的 API

何时*不*使用代理

有些情况下，代理并不是一个好的解决方案。避免在以下情况下使用代理：

- 高量、低利润的任务，例如基本的聊天支持
- 实时应用程序，例如即时欺诈检测
- 零错误系统，包括医疗或安全决策
- 需要确定性结果的高度监管行业

部署代理时管理风险的指南

让我们回顾一些有效代理架构的有用提示。

优先考虑有效代理架构的关键要素

保持代理架构简单。每个代理依赖于三个关键架构组件：

- **环境**：代理操作的数字空间
- **工具**：代理用于行动或观察的接口
- **系统提示**：指导代理操作的规则、目标和行为

关键要点：
从简单的代理动作开始。在确认代理可靠性能后，再增加任务复杂性。

评估您的部署计划

您需要评估您的部署计划，以评估和减轻与代理实施相关的风险。

如果风险等级是 . . .	实施以下响应策略 . . .
高风险且难以察觉	使用人工审查和多个验证层
高风险且明显	添加自动检查和监督机制
低风险	监控代理的行为、用户反馈和轻量级验证

最佳实践：

在开始使用代理时，请遵循以下最佳实践：

- 从**只读访问**工具和系统开始
- 对关键步骤添加**人工审批**
- 使用**分阶段部署**并进行监控
- 启用**全面日志记录**

接下来，了解如何负责任地实施代理。

负责任地实施代理

在实施代理时，请考虑一个分阶段的部署计划，包括以下步骤：

1. **验证概念证明**：尝试低风险、可逆的任务
2. **实施试点程序**：在监督下使用中等风险任务测试代理
3. **生产扩展**：仅在证明其安全性和性能后，扩大代理的使用

展望未来：将会有哪些改进？

智能体持续进化。期待未来的改进包括：

- 更一致的推理
- 更智能、更精简的架构
- 先进的监控和错误检测工具

但请记住：即使有改进，**深思熟虑的部署和风险分析仍然至关重要。**

现在，让我们回顾一下您所学到的关键点。

回顾

你现在知道：

- **AI代理**位于AI复杂性谱的最高端，擅长需要自主决策、适应和战略的任务。
  使用四步决策框架——任务模糊性、步骤灵活性、工具多样性和失败影响——来判断代理是否适合该任务。
- **避免使用代理**处理简单、可重复或高风险的任务，因为这些任务的错误代价高昂或可预测；工具的表现可能更好。
- **今天的代理**在可靠性、高计算成本方面存在困难，通常需要人类监督以避免产生幻觉或失误。
- **管理风险**的方法包括设定边界、使用日志、监测结果，并保持人类参与以进行监督。
- **有效的代理架构**包括模块化组件，如记忆、工具使用、规划策略和清晰的推理路径。

Author

Tenzin Migmar

贡献者

Faranak Heidari