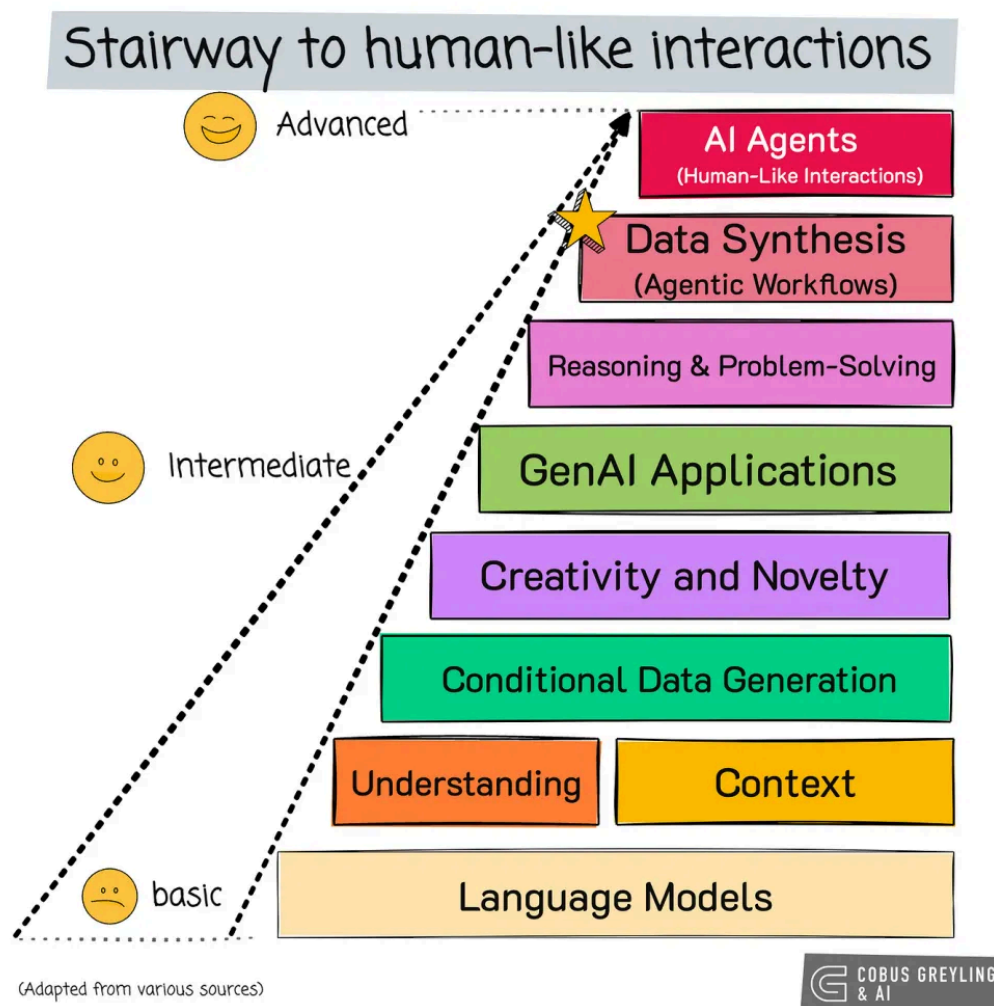


从AI Agent到Agentic Workflow：技术焦点的转变与未来发展

原创 南七无名氏 PyTorch研习社 2025年02月10日 08:01 安徽

当前，我们正处于一个重要的技术转折点。从大型语言模型（LLM）的诞生，到具有人类数字互动的AI Agent 的出现，技术进步为我们带来了巨大的变革。然而，在实际的商业化应用中，焦点正在发生变化，AI Agent 的关注逐渐转向了 Agentic Workflow 和数据合成。

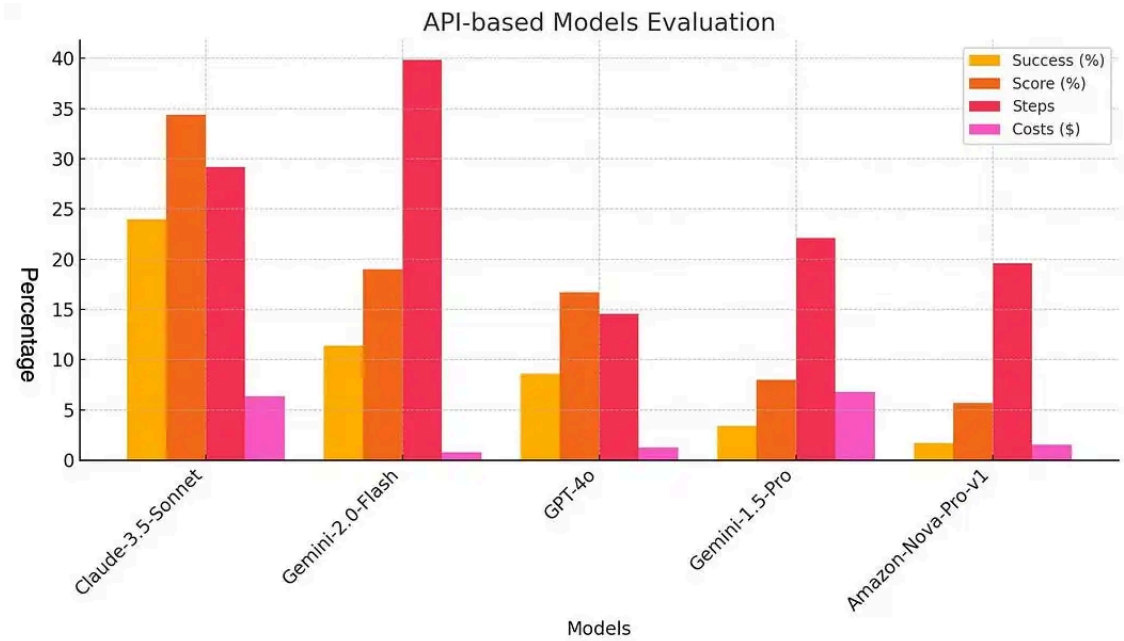


为什么焦点暂时从 AI Agent 转移？

像 Salesforce 和 Service 等公司曾大力推动 AI Agent 的应用，然而，AI Agent 面临的现实问题是，当前的技术准确性仍然未能达到预期标准。

尽管市场营销热炒和展示了多个令人印象深刻的 AI Agent 原型，但如果剥去这些外衣，AI Agent 的准确性仍不适合大规模生产环境使用。

例如，Claude AI Agent Computer Interface (ACI) 的性能，仅为人类的14%。



根据 TheAgentFactory 的数据显示，AI Agent 的表现主要集中在成本、步骤和成功率方面，成功率大约为 20%。这些数据直观地反映了 AI Agent 在实际应用中的瓶颈和现实困境。

随着 OpenAI Operator 的发布，计算机和网页浏览的准确率已达 30% 到 50%，但与人类的 70%+ 准确性相比，仍有明显差距。同时，研究表明，具备网页浏览功能的 AI Agent 容易受到恶意弹窗攻击，安全性存在隐患。

Benchmark type	Benchmark	Computer use (universal interface)		Web browsing agents	Human
		OpenAI CUA	Previous SOTA	Previous SOTA	
Computer use	OSWorld	38.1%	22.0%	-	72.4%
Browser use	WebArena	58.1%	36.2%	57.1%	78.2%
	WebVoyager	87.0%	56.0%	87.0%	-

AI Agent 执行任务的方式有两种：

- 一种是网页浏览器（如Webvoyager、OpenAI Operator等）；
- 另一种是通过操作系统的完整GUI（如Anthropic）。



这些方法利用 GUI 作为 AI Agent 的 API。

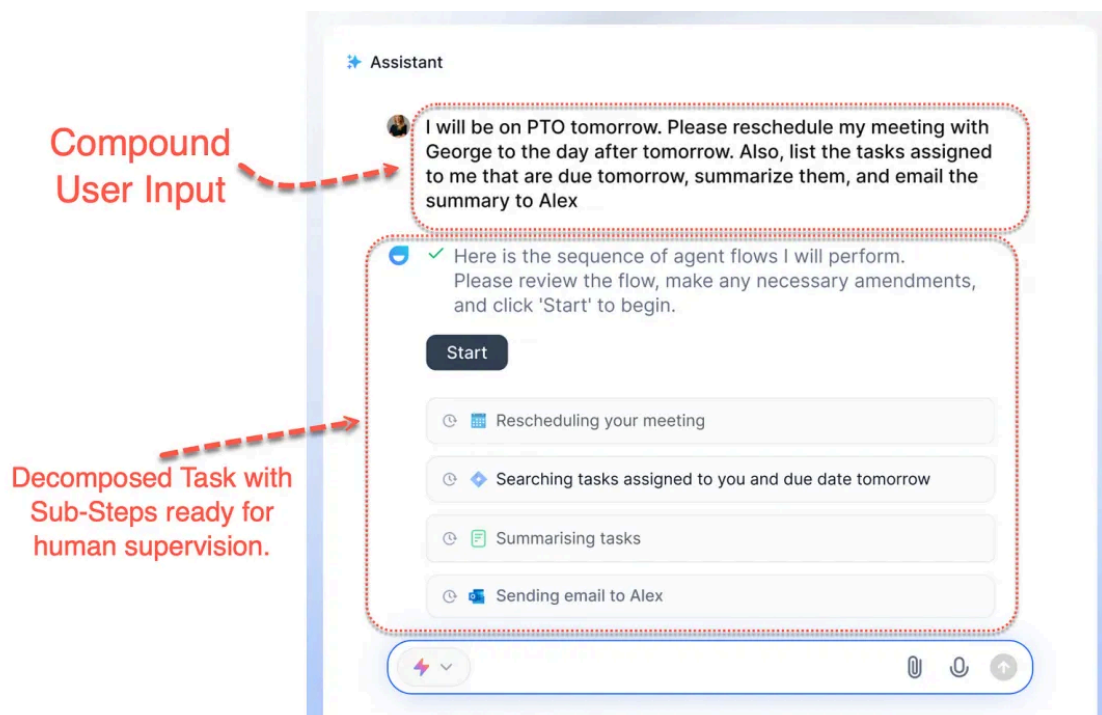
最初的方案是使用单独的 API，但由于每个 API 集成开发的开销，这种方式并不实际。而且，对于许多商业应用而言，并没有现成的 API。

为什么焦点转向 Agentic Workflow?

如今，大家普遍认为现代知识工作存在诸多问题，各种数据展示了这个现象。例如，有报告指出，员工花费30%的时间在搜索信息上。

知识工作者在解答复杂问题时面临的挑战也不容忽视，他们需要从多个文档中合成信息。

Agentic Workflow（如下图所示）能够进行推理，将复杂任务分解成更简单的子任务，并将这些任务按顺序链式执行。



通过执行这个顺序，像可观察性、可检查性和可发现性等元素被引入。

数据的合成将变得越来越重要。Agentic Workflows为知识工作者提供了一个例子，它将工作数据和资源合成，帮助工作者快速得出答案。

语言模型提供商不再仅提供模型，而是扩展到用户体验领域。ChatGPT 中的 Deep Research 并不是一个新模型，而是 ChatGPT 中的一个新型 Agentic 能力，它能够进行多步互联网研究，解决复杂任务。它能在十几分钟内完成一个人类需要花费数小时才能完成的任务。

这也是如何将不同来源的数据合成以回答用户问题的一个好例子。

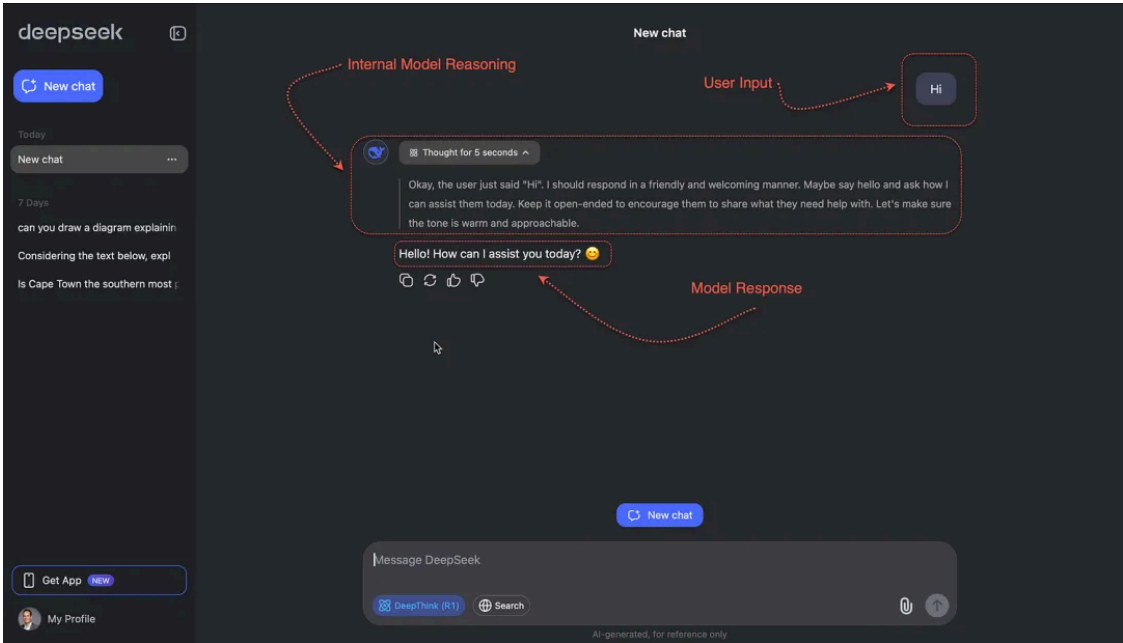
我认为这是 LlamaIndex 提出的一个概念——Agentic RAG，核心思想是将数据合成提供给“一个特定时刻的受众”变得越来越重要。

在未来几个月，个人 Agentic Workflow 和信息合成将成为焦点，某种程度上可以称之为桌面编排。



推理与问题解决

现代 AI 模型越来越多地将推理作为核心功能之一，使其能够通过将复杂问题分解为可管理的组件来解决这些问题。



这一转变的基础是创新的方法，即将问题分解成更小的子集，使模型能够系统地处理每个部分。

通过将推理作为一种内部机制，这些模型能够模拟类似人类的思维过程，增强它们提供准确和细致响应的能力。

分解策略不仅提高了解决问题的效率，还增强了得出结论过程的透明度。

因此，用户能够从更具可解释性的输出中受益，缩小了高级计算与易懂决策之间的鸿沟。

最初，用户需要在提示中包含推理特征，指示模型如何推理并分解复杂或复合任务。通过几次示范的方式，模型可以模仿并应用这些方法。

最后

企业必须将焦点从专注于某些特定工具或趋势——例如曾经自称为 RAG 公司、Prompt Engineering 实验场等，转移到解决现实世界的商业挑战上。

世界正以前所未有的速度向前发展，几乎每天都有新技术涌现，每一项都承诺能够颠覆各个行业。

然而，创新的真正衡量标准不在于掌握最新的技术，而在于将这些进步应用于创造实际的价值。

无论是改善客户体验、简化运营，还是解决社会需求，关键问题始终是：我们如何利用技术来提供有意义的解决方案？

通过采纳这种思维方式，企业能够为未来做好准备，并确保在不断变化的进步浪潮中保持相关性。

