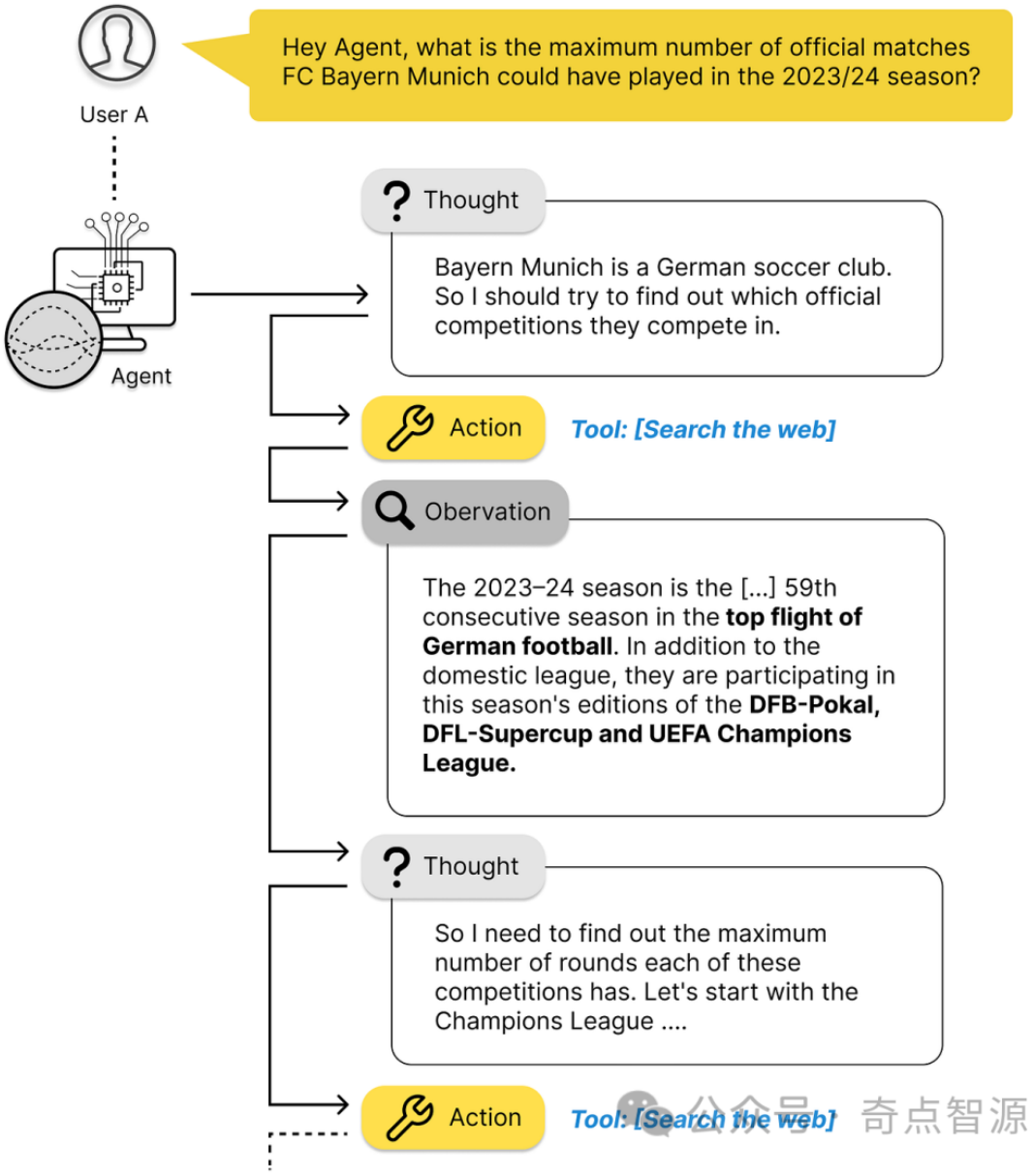
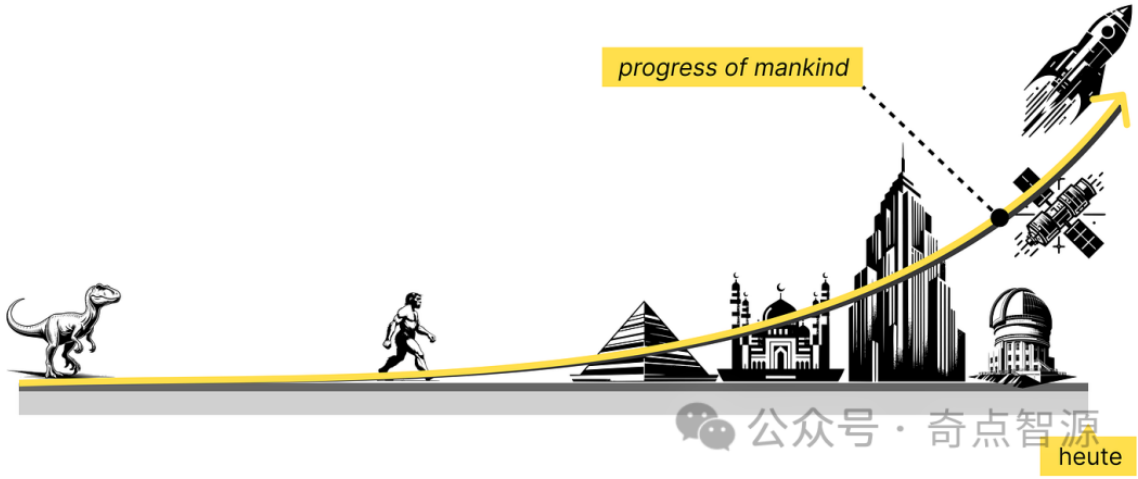
大语言模型智能体(LLM Agents)入门指南



“本文将探讨如何让大型语言模型（LLM）通过智能体（Agents）独立解决复杂任务。”

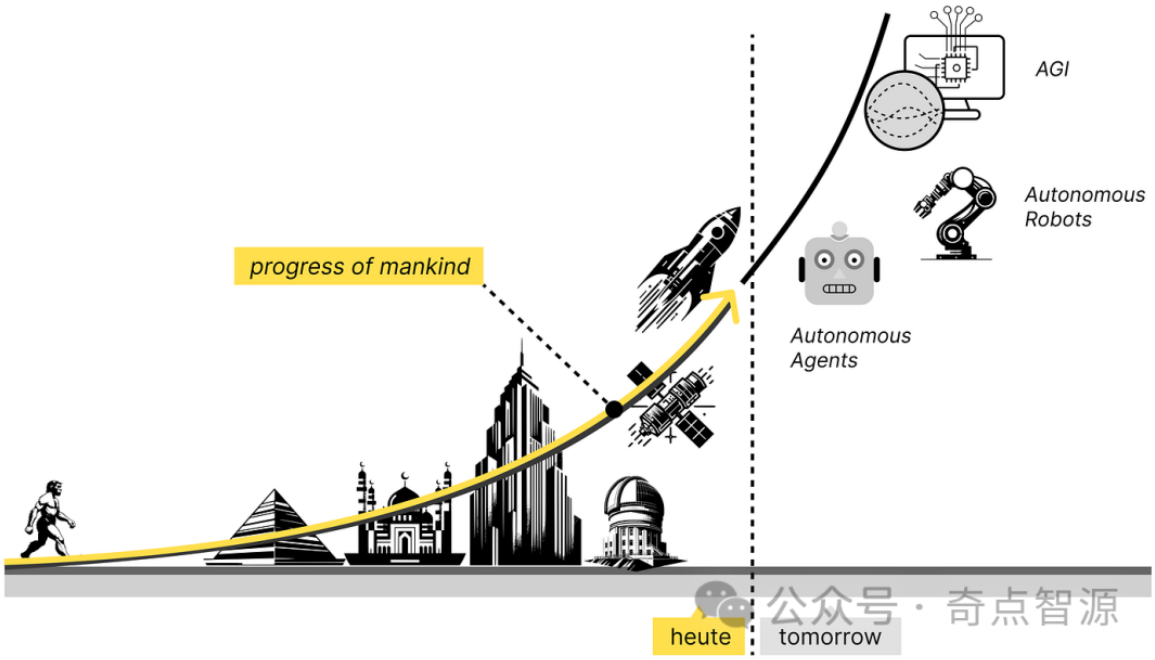
人类具备一项非凡的能力，那就是能够持续吸收信息、做出决策、采取行动、观察变化，并由此做出下一个决策。**我们的整个生命就是一个永无止境的观察、思考和行动的循环。**研究人员希望将这个概念应用到大型语言模型（LLM）中，使其能够不断做出新的决策，逐步接近具备解决复杂问题的能力。

人类通过将复杂问题分解为易于管理的小部分，并不断利用前人积累的知识，已经取得了长足的进步。现在我们已经到达了私营企业建造火箭、工厂完全由机器人操作的阶段，这个过程耗费了40亿年。



历史长河中的进步与创新

这并非终点，我们的经济虽然在不断地增长，而想要进一步加速进步只能依靠颠覆性的理念。**其中，生成式人工智能无疑是重要的一环，它不仅能生成新的内容，还能解释文本和观察结果并自主做出决策。**通过不断分析观察和做出决策，我们推动着进步，直到最终在火星上建造城市、实现永生，或是实现任何我们想要达到的目标。



现在以及未来的进步与创新

然而，今天我们还没有完全做到这一点，因为现有的每个人工智能模型都只模仿了人类智能的某个特定方面。例如，大型语言模型在理解和创作文本方面非常出色，其能力甚至超过了人类。但是，当涉及到简单的算术任务时，大型语言模型往往会遇到困难。

那么，如何让它们能够独立解决更复杂的问题呢？—— 其中一种方式是利用**智能体 (Agents)** 的概念。

01

什么是智能体

科幻电影和间谍电影中经常出现一种中央人工智能，它与主角交流，搜索互联网和各种秘密数据库，引导主角完成任务。例如电影《钢铁侠》中的贾维斯（J.A.R.V.I.S.）就是一个典型例子。

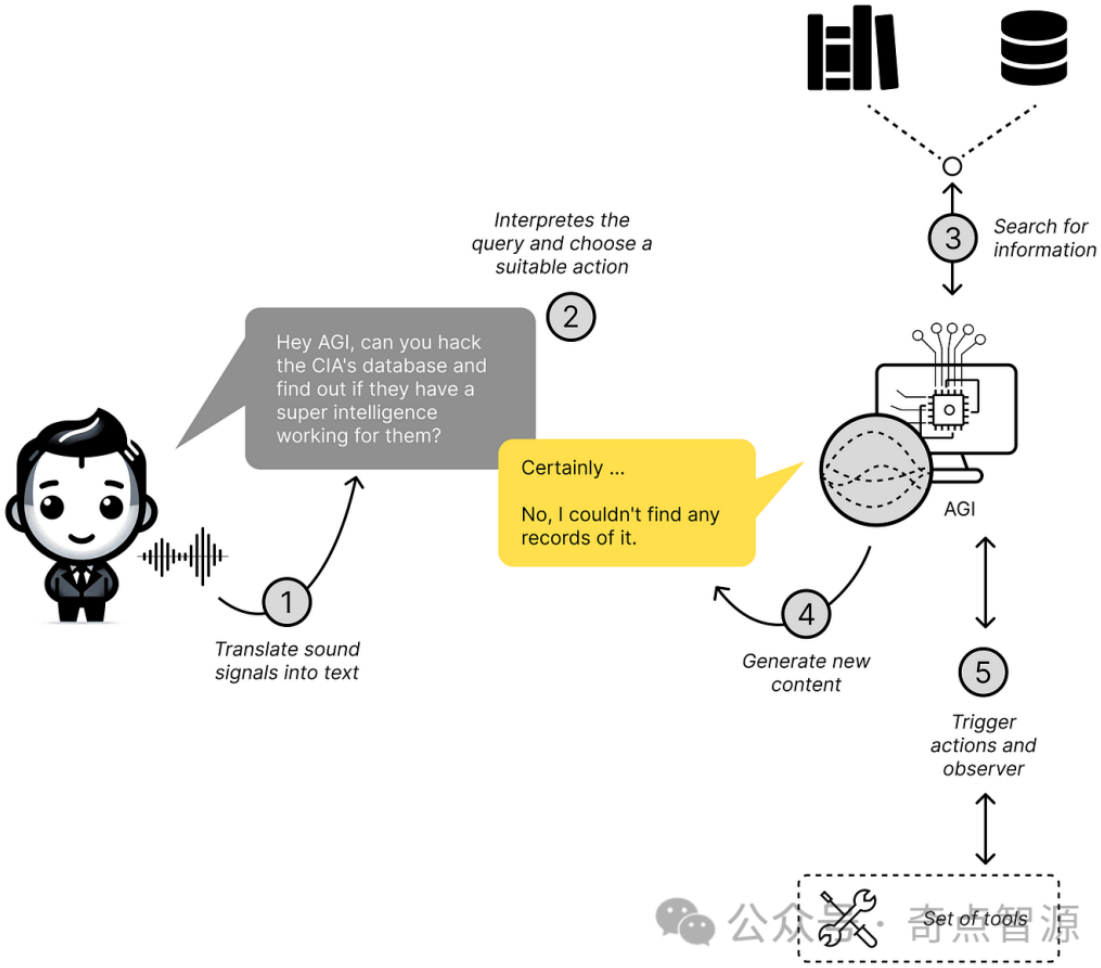
是什么让贾维斯如此特别？**钢铁侠甚至不需要告诉它如何解决问题，它会自己找到方法。**

这正是我们希望通过智能体实现的目标。

02

如何构建智能助手

核心毫无疑问是超级强大的大型语言模型，它能够理解问题、观察环境并基于此做出决策。除此之外，再加上一些将语音转换为文本的模型以及解释图像内容的模型，我们就拥有了构建自己的“贾维斯”所需的一切。



智能助手需要掌握的技能

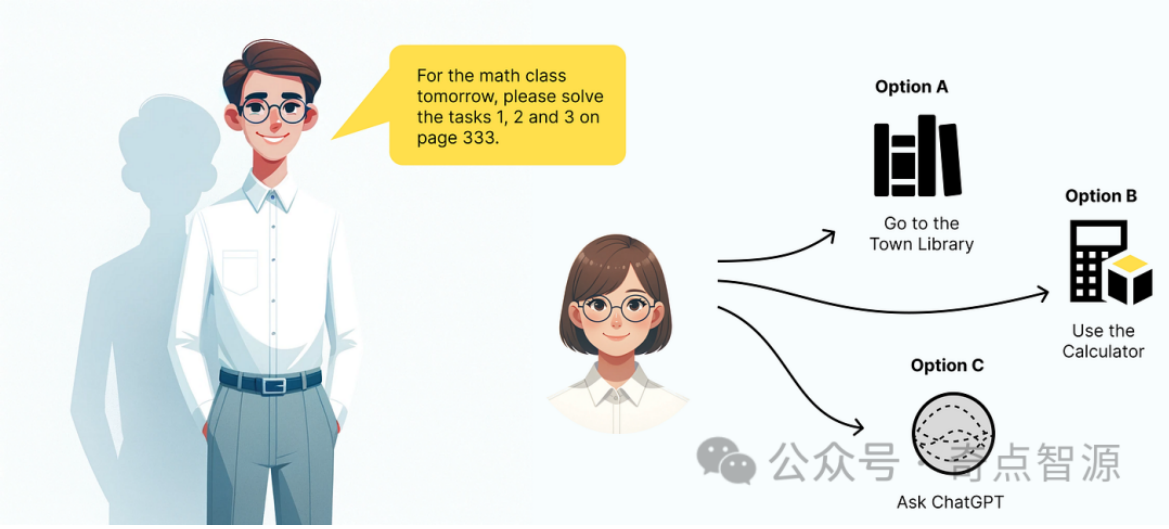
03

如何连接各项技能 - 智能体背后的理论

智能体是一种可访问一系列工具的组件。**其主要特征在于它能够做出明智的决策并利用适当的工具，直到找到足够好的答案**。

在大型语言模型应用方面，智能体的概念可能是最引人注目的进展。它让我们梦想着拥有一个能够自主控制流程、进行研究或通过找到生存概率最高的方案来拯救超级英雄的人工智能。

当我们应用智能体的概念时，我们不仅仅使用大型语言模型来回答问题，而是将其作为大脑，处理它看到的观察结果并决定接下来要做什么。我们人类一直在做着同样的事情：面对要解决的任务，寻找能够帮助我们尽可能轻松地完成任务的方法和工具。

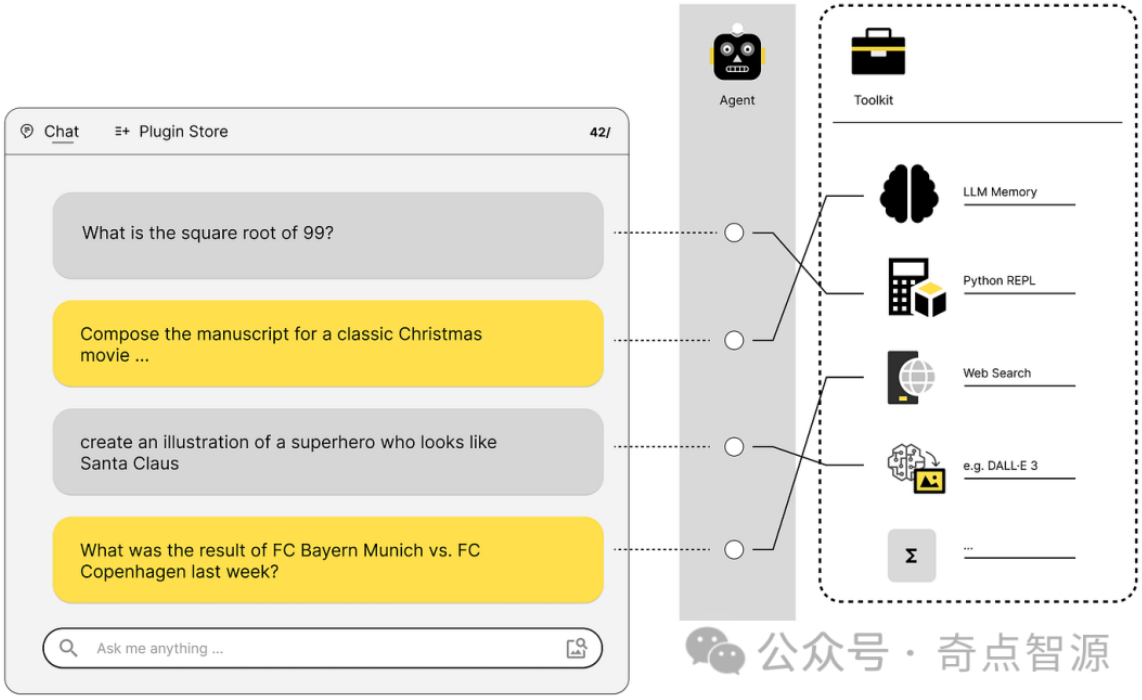


人类可以使用工具解决问题

通过应用这种概念和使用正确的工具，计算机也能够解决复杂的查询。例如，如果我们使用大型语言模型将文本转换为SQL，当执行生成的SQL时发生错误，LangChain 的 SQLAgent 不会轻易放弃，而是尝试解释错误并修正问题。

刚刚提到的SQL智能体使用了一套名为 SQLDatabaseToolkit 的 LangChain 工具。目前可用的工具并不局限于SQL。本文将在下面详细介绍 LangChain 已经提供的工具，不过请记住，工具可以是任何东西。之前提及大型语言模型在简单的分析计算方面有时表现不佳，因此解决这类问题需要的是一个能够识别分析任务并使用计算器解决问题的智能体。

OpenAI的ChatGPT已经做到了这一点。如果你通过Plus订阅发送请求，你会看到一个分析步骤，聊天机器人根据分析的结果决定如何响应请求。它会从一系列工具中进行选择，例如Python REPL用于解决分析任务，需要最新信息时使用网络搜索，用户要求创建图像时使用Dall-E3等等。



可选的工具以及使用场景

04

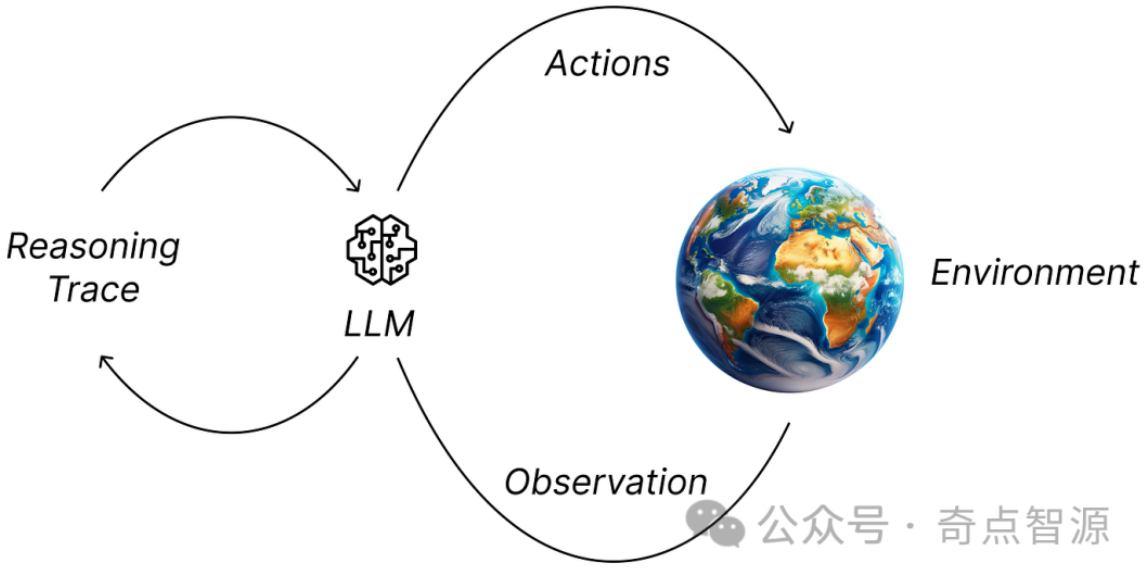
智能体的运作方式 - 思维链

人类的一大优势就是能够吸收大量的信息，过滤掉不重要的细节，并基于关键信息做出决策。我们通常首先将大问题分解为假设，然后尝试通过观察逐步支持或反驳这些假设。

我们使用“**思维链提示**”模拟这一过程，将多步骤问题分解为中间步骤。

它可以用3个简单步骤来描述：

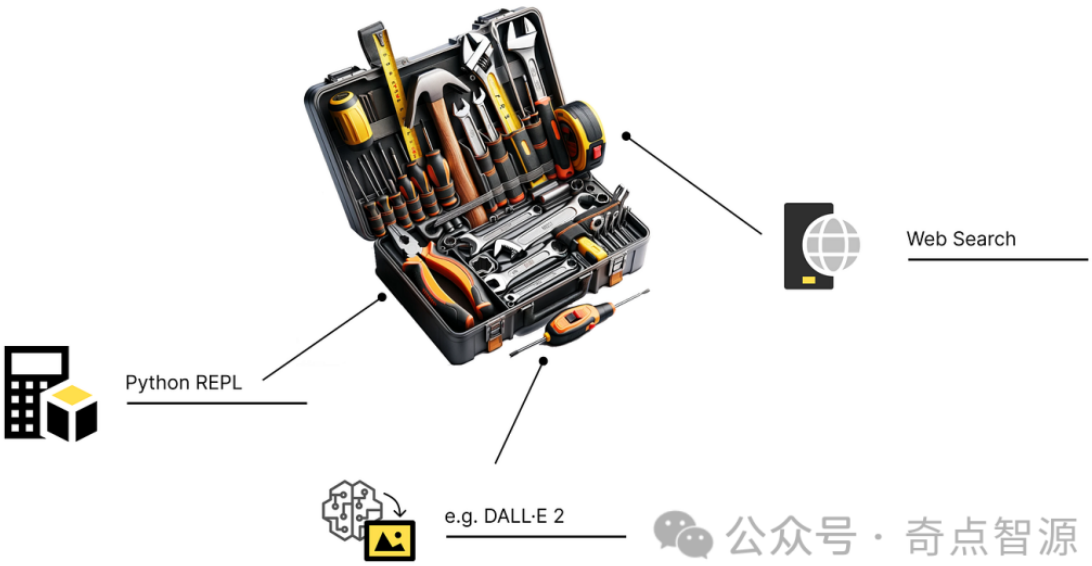
1. 我们发起一个行动 (Action)，大型语言模型观察来自选定环境的反馈。
2. 收集所有信息并利用这些信息来决定下一步采取什么合适的行动。
3. 如果我们反复执行这些步骤来解决更复杂的任务，我们会使用一种称为“推理追踪 (reasoning trace)”的方法，追踪过程中经历的步骤或阶段，以得出结论或解决方案。



思维链：智能体如何寻找解决方案

因此，如果要实现类似功能，首先需要一个潜在行动的集合。LangChain 以“工具箱”的形式来提供可能的功能组合。

这些工具箱提供各种函数或功能供大型语言模型根据当前任务进行选择。



工具箱

工具的描述方式需要让智能体清楚地知道每个工具的用途。这合乎常理，因为大型语言模型和智能体无法施展魔法，它们至少需要这些描述作为输入，才能决定使用哪些工具完成任务。

LangChain已经提供了一些内置工具，同时也允许你创建自己的工具集。

那么还剩下哪些内容有待了解呢？我们需要一个使用这些工具的 “执行者”，这就是智能体执行器 (Agent Executor)。

智能体执行器 - 智能体的幕后主脑

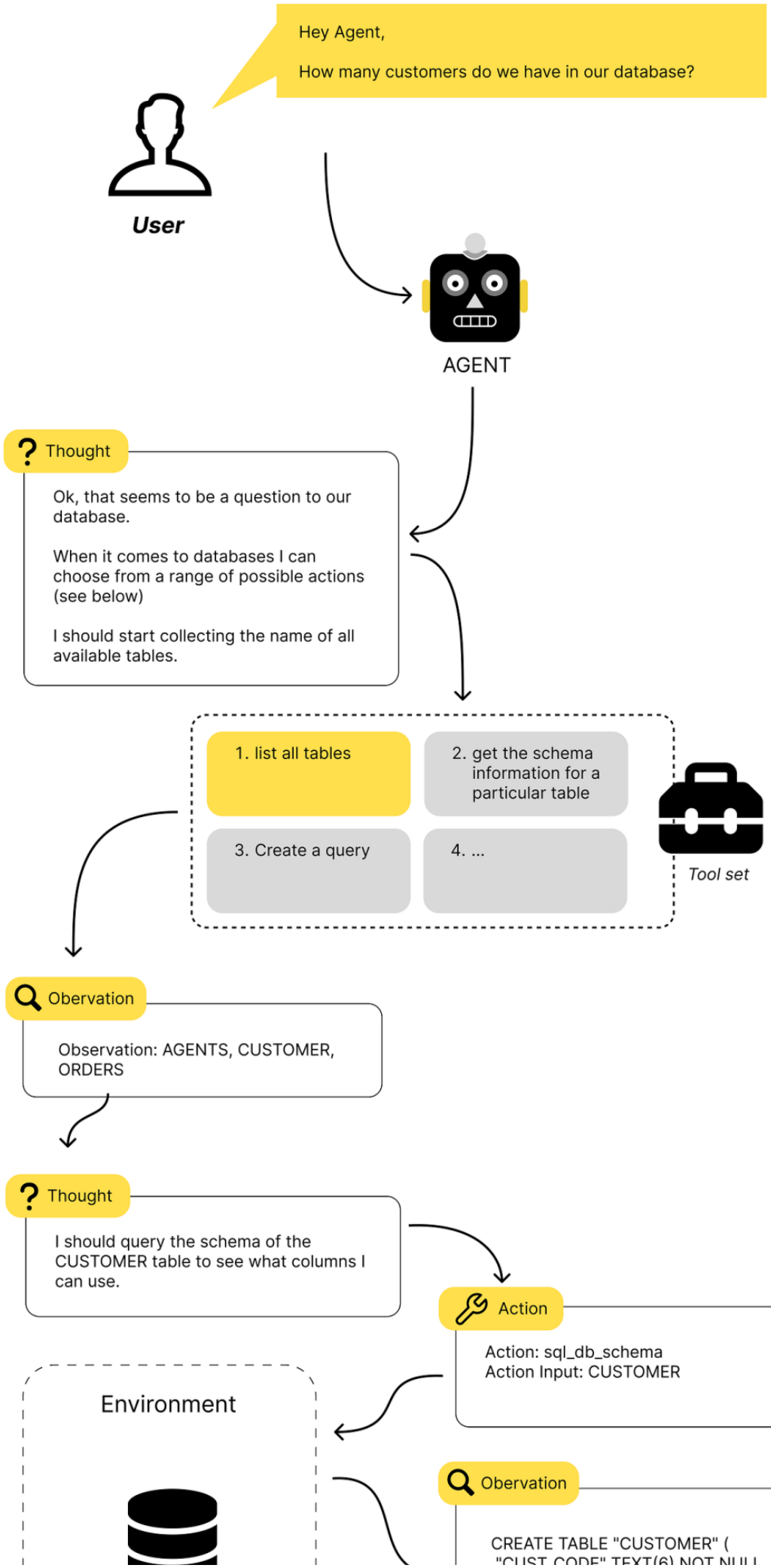
智能体能够连续使用正确的工具，持续观察结果，然后决定下一步需要哪些工具。这种迭代执行功能的操作由所谓的智能体执行器完成。

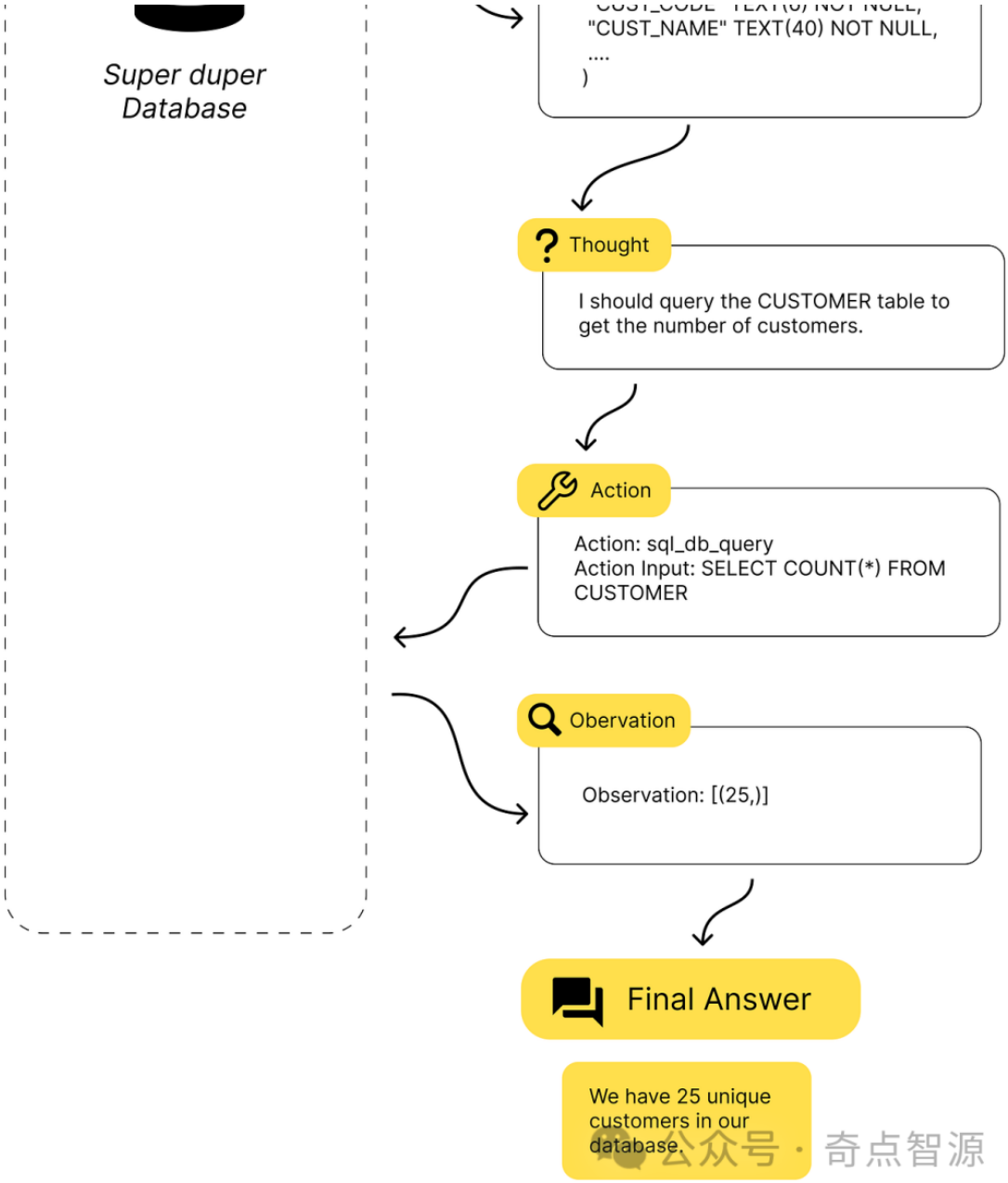
整个过程会反复进行，直到达到预定义的终止条件。下图展示了一个例子，使用 LangChain 中的 SQLDatabaseToolkit 来找到“我们的数据库中有多少客户”这个问题的答案。

这个例子中的数据库包含三个表格：

1. 客户：所有客户以及描述客户的数据
2. 订单：所有与客户表中的客户相关的订单
3. 代理：分配给不同客户的代理

让我们看看智能体如何尝试找到问题的答案。


Hey Agent,
How many customers do we have in our database?
User
AGENT
Thought
Ok, that seems to be a question to our database.
When it comes to databases I can choose from a range of possible actions (see below)
I should start collecting the name of all available tables.
1. list all tables
2. get the schema information for a particular table
3. Create a query
4. ...
Tool set
Obervation
Observation: AGENTS, CUSTOMER, ORDERS
Thought
I should query the schema of the CUSTOMER table to see what columns I can use.
Action
Action: sql_db_schema
Action Input: CUSTOMER
Environment
Obervation
CREATE TABLE "CUSTOMER" (



智能体如何找到解决方案

刚刚描述的智能体执行器实际上只是众多可选方案之一。你也可以选择其他智能体执行器，例如：

- Plan-and-execute Agent
- Baby AGI
- Auto GPT

每个智能体执行器都有自己的执行任务和做出决策的方法和模式。选择哪种运行时取决于具体任务的特定要求、决策过程的复杂性以及你希望智能体展现的自主性或智能程度。

总而言之：

- **工具**是智能体可以利用的各种功能、能力或行动，用来完成任务或解决问题。
- **这些工具必须清晰描述**，让智能体理解它们的功能以及如何有效使用它们。

- Langchain不仅提供一系列内置工具，还具有添加自定义工具的灵活性。
- 智能体执行器在管理工作流程、持续评估所用工具的有效性以及根据需要调整策略以实现预期结果方面发挥着关键作用。

06
总结

我们人类将复杂问题分解成更小的子任务和假设，并试图一步一步地证明或证伪它们，以逐步接近解决更大的难题。我们希望通过使用智能体的概念，用大型语言模型模拟这种行为。**通过为智能体提供正确的工具，智能体能够自主决定下一步采取什么行动，以更接近最终的解决方案。**

这一切非常令人兴奋，因为如果智能体真的变得足够智能，能够模仿人类的独创性和研究技能，我们就能在这些领域取得巨大的进步。在发明新事物和探索世界奥秘方面，人类不再会成为瓶颈。