

2024

中国AI Agent 行业研究报告

大模型时代的“APP”，探索新一代人机交互及协作范式

出品机构：甲子光年智库

智库院长：宋涛

撰写分析师：刘瑶、胡博文

发布时间：2024.4

目录

CONTENTS

Part 01 前世今生：科幻憧憬、学术概念与商业尝试

Part 02 奇点已至：让每个人掌握AI的力量

Part 03 百家争鸣：属于大模型时代的APP繁荣

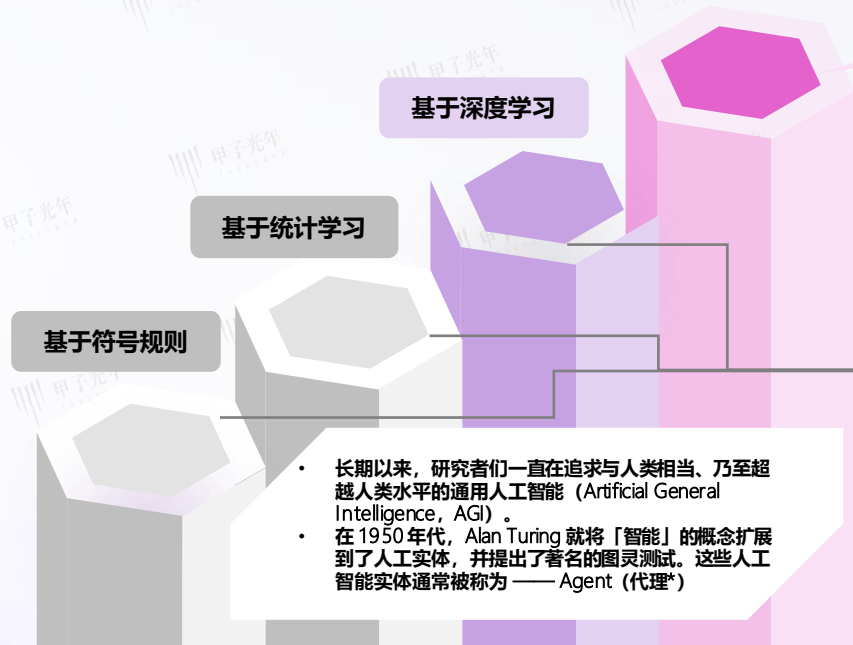
Part 04 时代先驱：当下商业实践值得关注的里程碑

Part 05 潜力无限：来自于数据、算法、算力的飞轮效应

AI Agent的发展历程梳理：大模型赋予了AI Agent核心改变

- Agent (代理) 一概念起源于哲学，描述了一种拥有欲望、信念、意图以及采取行动能力的实体。在人工智能领域，这一术语被赋予了一层新的含义：**具有自主性、反应性、交互性等特征的智能“代理”**。大型语言模型 (LLMs) 的出现为智能代理的进一步发展带来了希望。

AI Agent的发展历程简述



*Agent 术语的中文译名并未形成共识，有学者将其翻译为智能体、行为体、代理或智能代理，目前行业出现的“代理”和“智能代理”均指代 Agent，后由于2023年Open AI引爆AIGC领域，一般称为AI Agent。

数据来源：公开资料，专家访谈，甲子光年智库整理

基于大模型

LLM给AI Agent底层提供了一个**突破性技术方案**：LLM带来了深度学习新范式，思维链和强大的自然语言理解能力有望让Agent 具备强大的学习能力和迁移能力，从而让**创建广泛应用且实用的Agent**成为可能

LLM的框架优势：过去等强化学习基于深度学习框架可让Agent学到技能，但Agent的泛化性较差，往往用于非常窄的特定领域，例如用在游戏或低维层面的控制或计划，标志性应用是围棋领域的AlphaGo。

过往的工作主要集中在增强代理的特定能力，如符号推理或对特定任务的掌握（国际象棋、围棋等）。这些研究更加注重算法设计和训练策略，而忽视了模型固有的通用能力的发展，如知识记忆、长期规划、有效泛化和高效互动等。事实证明，**增强模型固有能力和推动智能代理进一步发展的关键因素。**

过往的AI Agent类型：

- **符号型智能体**：采用逻辑规则和符号表示来封装知识和促进推理过程，如1980年前后，出现的医学诊断专家系统，模拟心理治疗程序等；
- **反映型智能体**：关注智能体与其环境之间的交互，强调快速和实时响应，缺乏复杂决策和规划能力；
- **基于强化学习的智能体**：关注如何让智能体通过与环境的交互进行学习。
- **基于迁移学习和元学习的智能体**：使智能体从少量样本中迅速推理出金刀舞的最优策略。

LLM是Agent能力的增效器，交互协作程度是Agent能力的扩展器

- 当下大模型的参数量提升AI Agent的理解力和泛化能力，使其能更好地处理多种任务和上下文信息。这增强了AI代理的自然语言处理能力，从而提供更个性化、连贯的交互体验，是当下Agent的构建关键。

核心特征

大模型时代的AI Agent = LLM × (规划+记忆+工具+行动)

LLM是核心控制器，构建核心能力

提升AI Agent的理解力和泛化能力，使其能更好地处理多种任务和上下文信息。这增强了AI代理的自然语言处理能力，从而提供更个性化、连贯的交互体验。

架构解析

Agent基于LLM的组件，和交互两个层面



人类用户

通过用户接口、命令指示等方式与Agent形成交互，是Agent的监督者、合作者和决策者

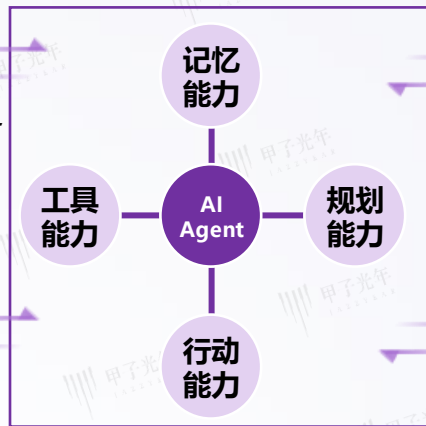


外界环境

Agent所处的环境（可能包括虚拟及物理世界），外界环境可以与Agent形成交互

人机交互

感知反馈



群体协作

设计开发



Agents

其他Agent，多Agent可以形成协作，结合相关任务结果形成群体智能

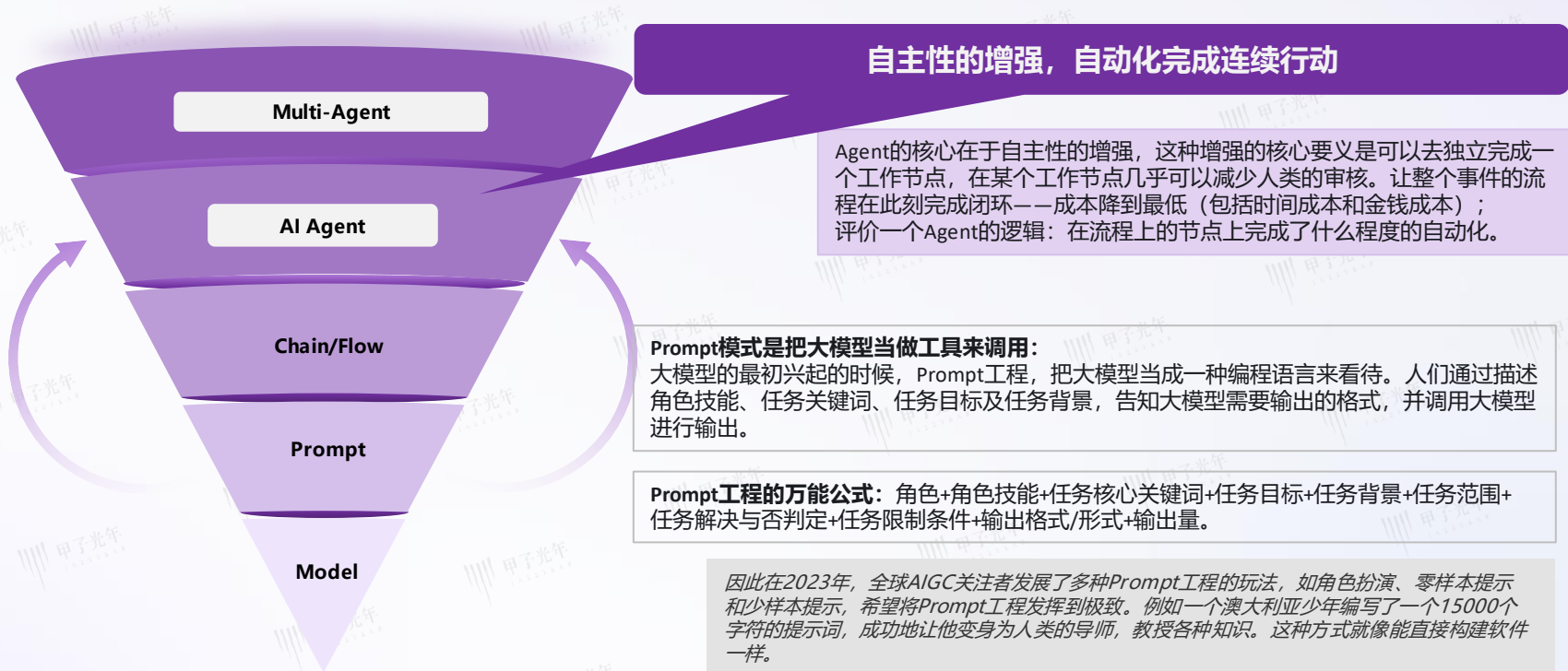


系统开发者

Agent的开发者，对Agent的相关能力，设计Agent的相关组件能力

当下的AI Agent可以看作LLM技术下Prompt工程的进化

- AI Agent是Prompt工程的一种升级，Agent的核心在于自主性的增强，可有效完成某一个工作点或工作单元，尽量减少人的干预；
- 评价一个AI Agent的核心逻辑：在流程上的节点上完成了什么程度的自动化。

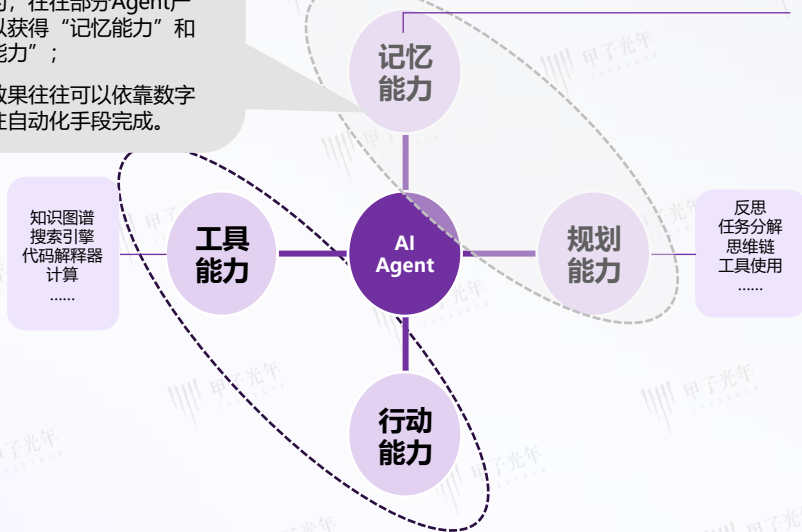


【记忆】和【规划】是学术概念下的关键功能点，商业概念也将逐步落地

- ❑ 从学术概念来看【记忆】和【规划】对于AI Agent学术概念上的完整性非常关键，但受限与市场发展早期，在实际的商业产品落地中【记忆】和【规划】能力未必能完全呈现
- ❑ 理解这点就能接受在当下市场环境下AI Agent产品功能的不完整，并且对Agent的形态持续保持关注和期待。

2024年仅仅距离AIGC全面爆发的2023年才过去1年，受限与算力、模型、数据等多种因素制约，往往部分Agent产品会难以获得“记忆能力”和“规划能力”；

产品的效果往往可以依靠数字化及过往自动化手段完成。



图：人类记忆的分类



表：人类记忆与AI Agent记忆的映射

人类记忆类型	Agent映射	例子
感觉记忆	学习原始输入的嵌入表示，包括文本、图像或其他形式，短暂保留感觉印象。	看一张图片，然后在图片消失后能够在脑海中回想起它的视觉印象。
短期记忆	上下文学习（比如直接写入prompt中的信息），处理复杂任务的临时存储空间，受有限的上下文长度限制。	在进行心算时记住几个数字，但短期记忆是有限的，只能暂时保持几个项目。
长期记忆	在查询时Agent可以关注的外部向量存储，具有快速检索和基本无限的存储容量。	学会骑自行车后，多年来再次骑起来仍能掌握这项技能，这要归功于长期记忆的持久存储。