

AI Agent (或者LLM Agent) 深度讲解

AI Agent ([eɪdʒənt], 代理人) 是一种能够感知环境、进行决策和执行动作的智能实体。它通常指的是在人工智能领域中，能够自主运作并完成特定任务的计算实体或程序。这些Agent可以通过传感器感知周围环境，并根据感知到的信息做出决策，然后通过执行器采取行动。

AI Agent的核心特点包括：

- 1) **自主性**: AI Agent具有独立思考和行动的能力，能够在没有人类直接指导下完成任务。
- 2) **交互性**: AI Agent能够与环境或其他Agent进行交互，这通常用于游戏、对话系统、推荐系统等场景。
- 3) **目的性**: AI Agent设计有明确的目标或意图，它们的行为是为了实现这些目标。
- 4) **适应性**: AI Agent能够根据环境的变化调整自己的行为，以适应新的情境。
- 5) **进化性**: 随着技术的发展，AI Agent的功能和智能水平也在不断提升。

总的来说，AI Agent的应用非常广泛，从简单的自动化任务到复杂的决策支持系统，都可以使用AI Agent来实现。例如，在工业自动化中，AI Agent可以监控机器状态并预测维护需求；在金融服务中，AI Agent可以帮助分析市场趋势并提供投资建议；在医疗领域，AI Agent可以辅助医生进行诊断和治疗规划。随着人工智能技术的不断进步，AI Agent的能力和应用范围将进一步扩大。

AI Agent在多个领域有应用，具体如下：

- 1) 医疗领域：AI Agent可以帮助医生分析病例，提供诊断建议，甚至辅助进行手术规划。例如，通过分析大量的医疗数据，AI Agent能够识别疾病模式，从而辅助医生做出更准确的诊断。
- 2) 金融领域：在金融行业中，AI Agent可以用于风险评估、欺诈检测、投资策略制定等。它们通过分析市场数据和用户行为，为投资者提供个性化的投资建议。
- 3) 教育领域：AI Agent可以个性化地适应学生的学习进度和风格，提供定制化的学习资源和反馈，从而提高教育质量和效率。
- 4) 零售与电子商务领域：在零售行业，AI Agent可以通过分析消费者的购物习惯和偏好，提供个性化的购物体验和推荐。例如，智能语音助手可以协助用户在线购物，提供客服支持。
- 5) 自动驾驶领域：AI Agent在自动驾驶汽车中扮演着核心角色，它们能够处理来自车辆传感器的数据，做出驾驶决策，确保行车安全。
- 6) 智能家居领域：在智能家居领域，AI Agent可以控制家庭设备，学习用户的习惯，自动调整家居环境，提高生活的便利性和舒适度。

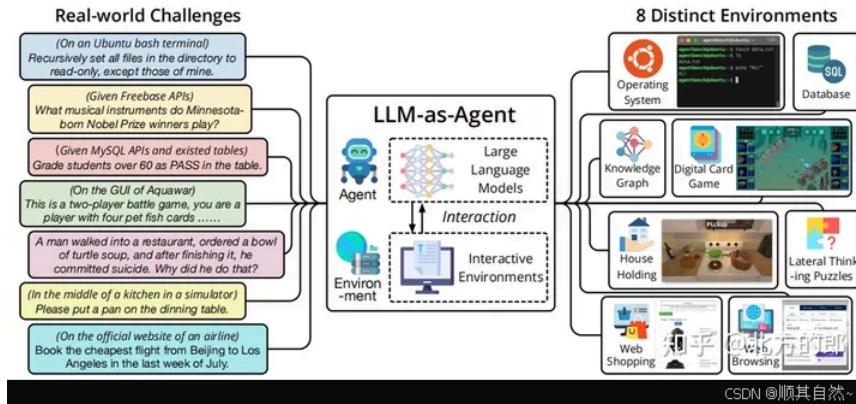
此外，AI Agent的应用不仅限于上述领域，随着技术的进步，它们有望在更多领域实现落地应用，如法律、制造业、农业等。AI Agent的智能交互、自主学习和灵活适应等特点，使其成为推动各行各业数字化转型的重要力量。

前一段时间，比尔·盖茨在他的博客上发表了：《AI is about to completely change how you use computers》比尔·盖茨在这篇文章中探讨了AI Agent对我们未来生活的影响。他谈到了AI Agent在医疗保健、教育、生产力、娱乐和购物等领域的作用。这些Agent将为人们提供更个性化的服务，帮助解决各种问题并提供支持，从辅助医生和教师工作到处理日常任务，甚至影响我们与朋友和家人的互动方式。AI Agent正在以各种方式迅速进入我们的生活，将在未来几年内彻底改变我们的生活方式。

1、什么是AI Agent (LLM Agent)

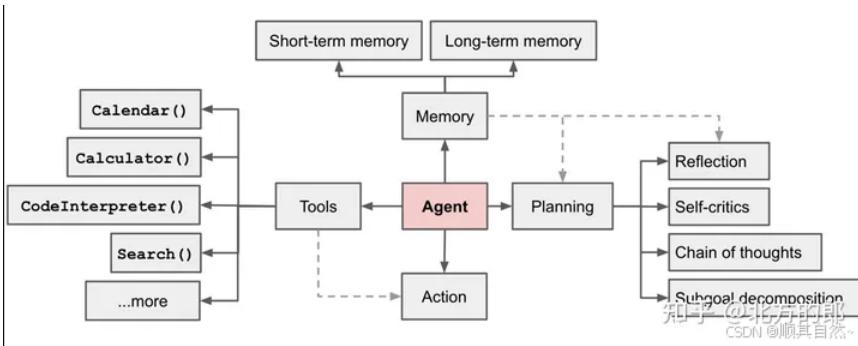
1.1、AI Agent 的定义

AI Agent是一种超越简单文本生成的人工智能系统。它使用大型语言模型 (LLM) 作为其核心计算引擎，使其能够进行对话、执行任务、推理并展现一定程度的自主性。简而言之，Agent是一个具有复杂推理能力、记忆和执行任务手段的系统。



1.2、AI Agent 的主要组成部分

在LLM赋能的自主agent系统中(LLM Agent)，LLM充当agent大脑的角色，并与若干关键组件协作。



规划 (planning)

- 1) 子目标分解: agent将大任务拆分为更小的可管理的子目标, 使得可以有效处理复杂任务。
- 2) 反思与完善: agent对历史动作可以自我批评和自我反思, 从错误中学习并在后续步骤里完善, 从而改善最终结果的质量。

记忆 (Memory)

- 1) 短期记忆: 上下文学习 (对话上下文) 即是利用模型的短期记忆学习。
- 2) 长期记忆: 为agent提供保留和召回长期信息的能力, 通常利用外部向量存储和检索实现。

工具使用 (tool use)

对模型权重丢失的信息, agent学习调用外部API获取额外信息, 包括当前信息、代码执行能力、专有信息源的访问等等

行动 (Action)

行动模块是智能体实际执行决定或响应的部分。面对不同的任务, 智能体系统有一个完整的行动策略集。在决策时可以选择需要执行的行动, 比如广为熟知的记忆检索、推理、学习、编程等。

2、AI Agent的意义

2.1、人机协同模式

基于大模型的Agent不仅可以让每个人都有增强能力的专属智能助理, 还将改变人机协同的模式, 带来更为广泛的人机融合。生成式AI的智能革命演化至今, 从人机协同呈现了三种模式:

(1) 嵌入 (embedding) 模式

用户通过与AI进行语言交流, 使用提示词来设定目标, 然后AI协助用户完成这些目标, 比如普通用户向生成式AI输入提示词创作小说、音乐作品、3D内容等。在这种模式下, AI的作用相当于执行命令的工具, 而人类担任决策者和指挥者的角色。

(2) 副驾驶 (Copilot) 模式

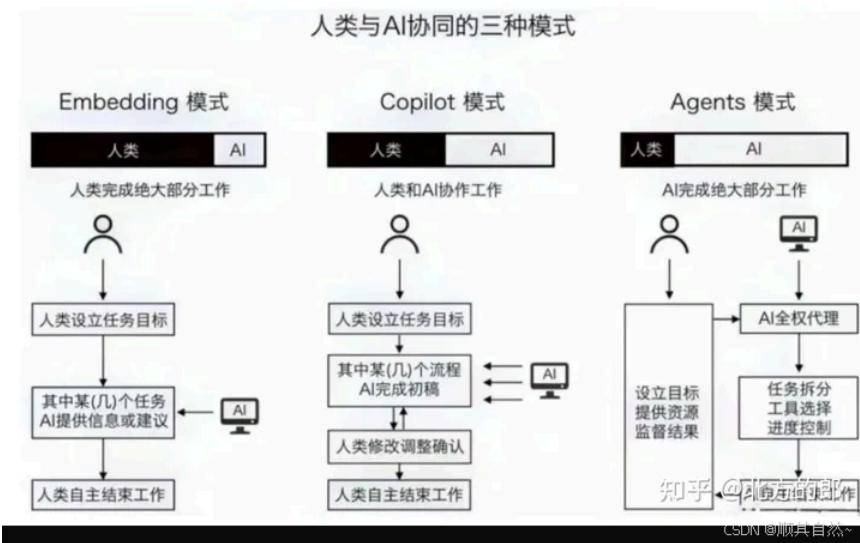
在这种模式下, 人类和AI更像是合作伙伴, 共同参与到工作流程中, 各自发挥作用。AI介入到工作流程中, 从提供建议到协助完成流程的各个阶段。例如, 在软件开发中, AI可以为程序员编写代码、检测错误或优化性能提供帮助。人类和AI在这个过程中共同工作, 互补彼此的能力。AI更像是一个知识丰富的合作伙伴, 而非单纯的工具。

实际上, 2021年微软在GitHub首次引入了Copilot (副驾驶) 的概念。GitHub Copilot是一个辅助开发人员编写代码的AI服务。2023年5月, 微软在大模型的加持下, Copilot迎来全面升级, 推出Dynamics 365 Copilot、Microsoft 365 Copilot和Power Platform Copilot等, 并提出“Copilot是一种全新的工作方式”的理念。工作如此, 生活也同样需要“Copilot”, “出门问问”创始人李志飞认为大模型的最好工作, 是做人类的“Copilot”。

(3) 智能体 (Agent) 模式

人类设定目标和提供必要的资源 (例如计算能力), 然后AI独立地承担大部分工作, 最后人类监督进程以及评估最终结果。这种模式下, AI充分体现了智能体的互动性、自主性和适应性特征, 接近于独立的行动者, 而人类则更多地扮演监督者和评估者的角色。

人类与AI协同的三种模式



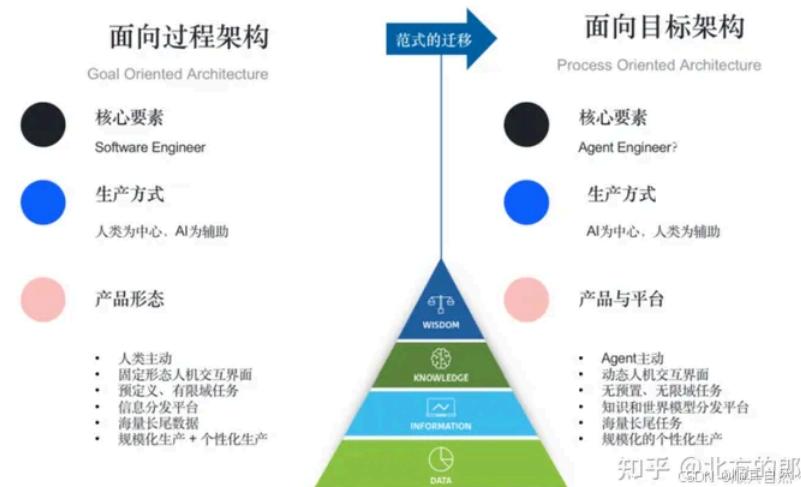
人类与AI协同的三种方式

从前文对智能体记忆、规划、行动和使用工具四个主要模块的功能分析来看，智能体模式相较于嵌入模式、副驾驶模式无疑更为高效，或将成为未来人机协同的主要模式。

基于Agent的人机协同模式，每个普通个体都有可能成为超级个体。超级个体是拥有自己的AI团队与自动化任务工作流，基于Agent与其他超级个体建立更为智能化与自动化的协作关系。现在业内不乏一人公司、超级个体的积极探索。

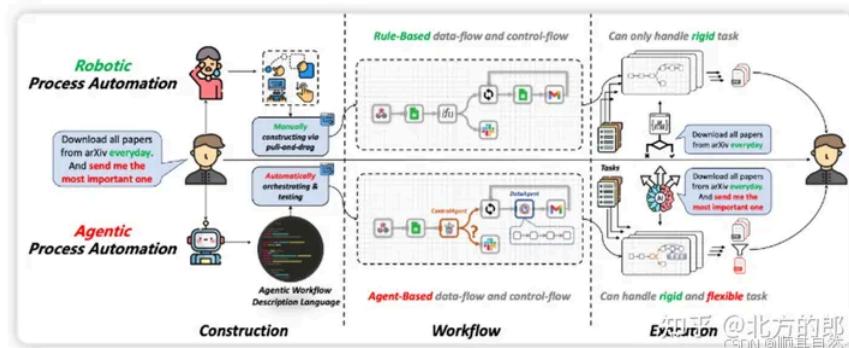
2.2、AI Agent与软件开发

AI Agent将使软件架构的范式从面向过程迁移到面向目标。现有的软件（包括APP）通过一系列预定义的指令、逻辑、规则和启发式算法将流程固定下来，以满足软件运行结果符合用户的预期，即用户按照指令逻辑一步一步操作达成目标。这样一种面向过程的软件架构具有高可靠性、确定性。但是，这种面向目标的架构只能应用于垂直领域，而无法普遍应用到所有领域，因此标准化和定制化之间如何平衡也成为SaaS行业面对的难题之一。



软件架构范式迁移

AI Agent范式将原本由人类主导的功能开发，逐渐迁移为以AI为主要驱动力。以**大模型为技术基础设施，Agent为核心产品形态**，把传统软件预定义的指令、逻辑、规则和启发式算法的任务层级演变成目标导向的智能体自主生成。这样一来，原本的架构只能解决有限范围的任务，未来的架构则可以解决无限域的任务。未来的软件生态，不仅是最上层与所有人交互的媒介是Agent，整个产业的发展，无论是底层技术，商业模式，中间组件，甚至是人们的生活习惯和行为都会围绕Agent来改变，这就是Agent-Centric时代的开启。



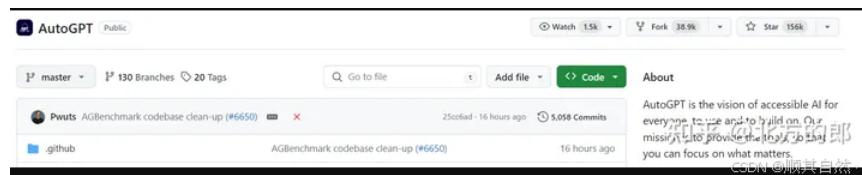
RPA范式 (Robotic Process Automation) 与APA范式 (Agentic Process Automation) 的比较

3、常见LLM Agent框架或者应用

3.1、AutoGPT

可以说是LLM Agent的鼻祖了。Auto-GPT是一个实验性的开源应用程序，展示了GPT-4语言模型的能力。这个程序由GPT-4驱动，将LLM“思想”连接在一起，以自主地实现您设置的任何目标。作为GPT-4完全自主运行的最早示例之一，Auto-GPT突破了人工智能的极限，将AI进程推向了新高度 -- 自主人工智能。

<https://github.com/Significant-Gravitas/AutoGPT>



The screenshot shows the GitHub repository page for AutoGPT. It displays the repository's name, a brief description, and various commit history and statistics.

AutoGPT: build & use AI agents

[AutoGPT](#) 50759 members [Follow @Auto_GPT](#) License MIT

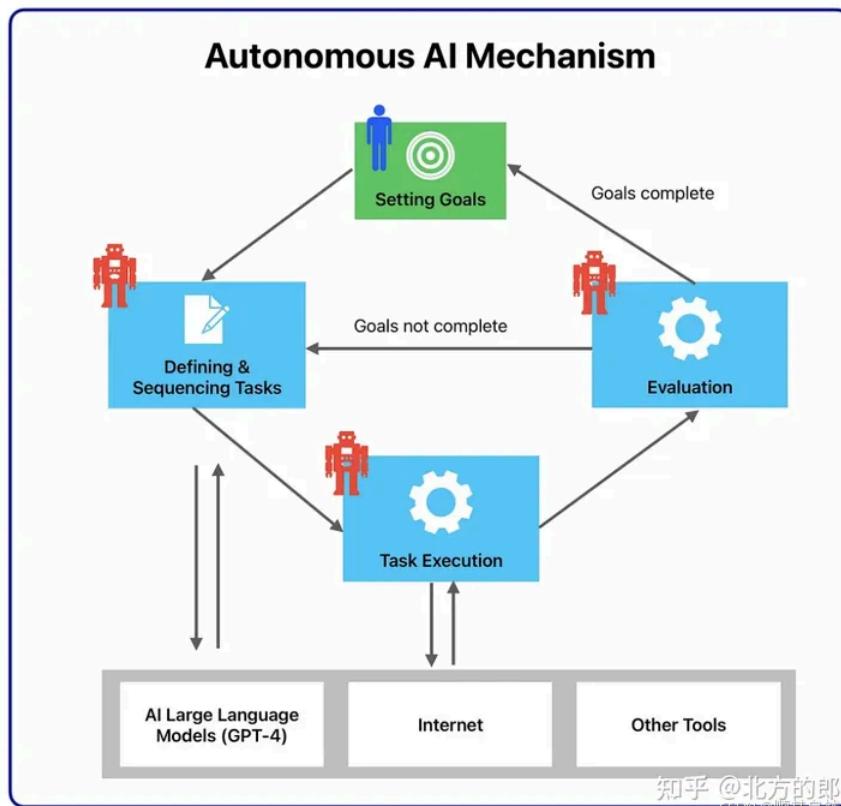
AutoGPT is the vision of the power of AI accessible to everyone, to use and to build on. Our mission is to provide the tools, so that you can focus on what matters:

- 🏗️ Building - Lay the foundation for something amazing.
- 🧪 Testing - Fine-tune your agent to perfection.
- 💖 Delegating - Let AI work for you, and have your ideas come to life.

Be part of the revolution! AutoGPT is here to stay, at the forefront of AI innovation.

[Documentation](#) | [Contributing](#) | [Build your own Agent - Quickstart](#)

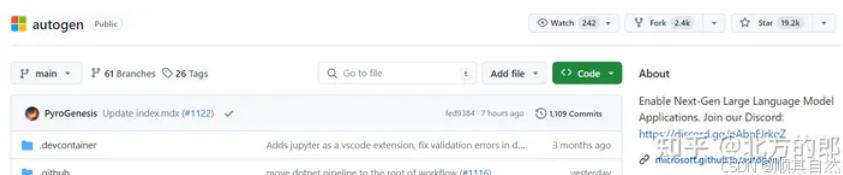
知乎 @北方的郎
CSDN 随顺其自然



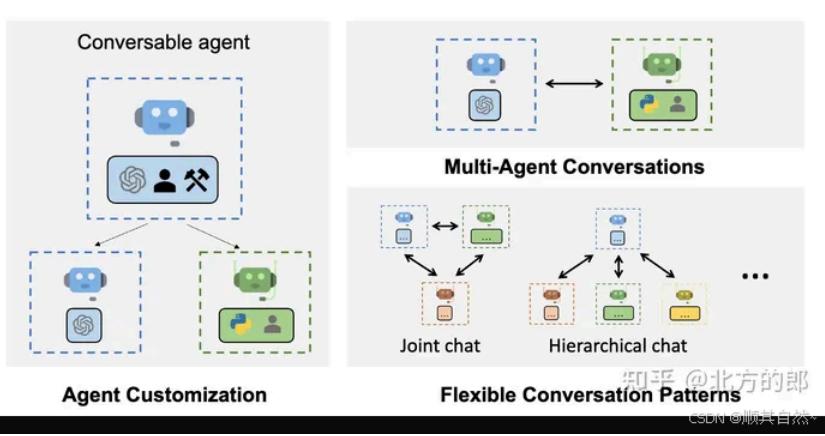
3.2、AutoGen

微软发布的AutoGen agent是可定制的、可对话的，并能以各种模式运行，这些模式采用LLM、人类输入和工具的组合。使用AutoGen，开发人员还可以灵活定义agent交互行为。自然语言和计算机代码都可用于为不同的应用编程灵活的对话模式。AutoGen可作为一个通用框架，构建具有不同复杂性和LLM能力的各种应用。实证研究证明了该框架在许多样本应用中的有效性，应用领域包括数学、编码、问答、运筹学、在线决策、娱乐等。

<https://github.com/microsoft/autogen>



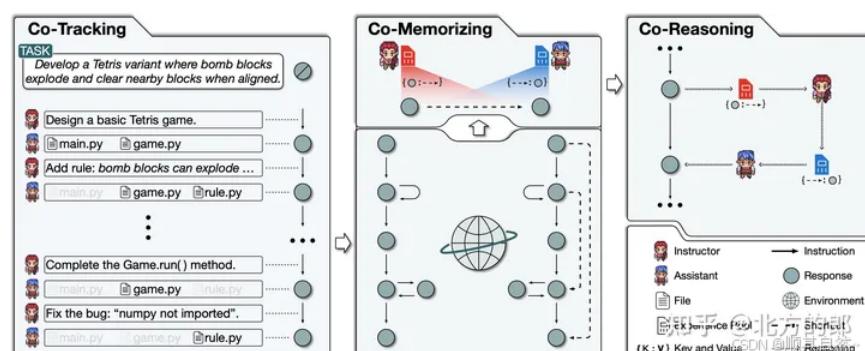
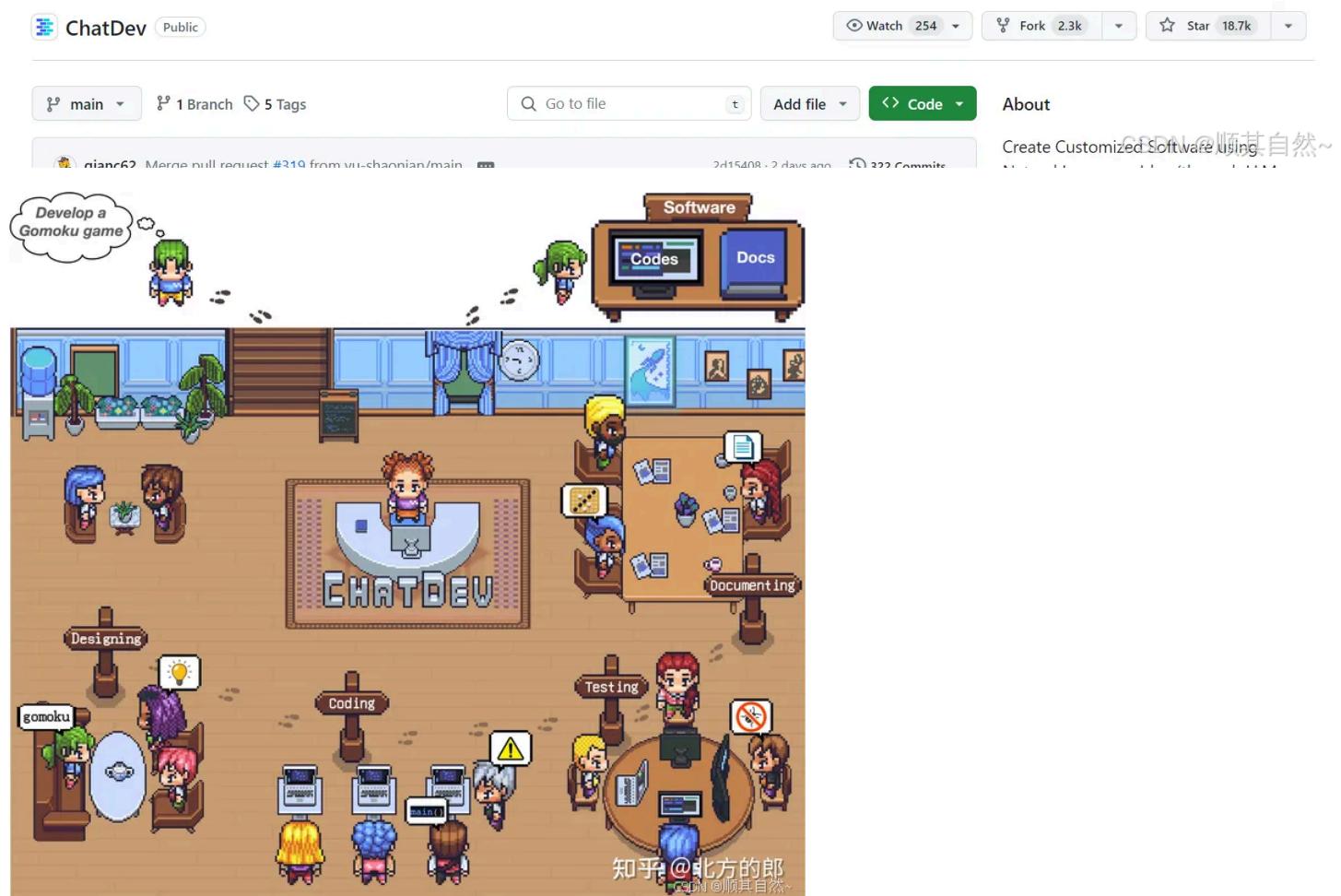
The screenshot shows the GitHub repository page for AutoGen. It displays the repository's name, a brief description, and various commit history and statistics.



3.3. ChatDev

清华大学 NLP 实验室联合面壁智能等科研机构研发的一个大模型驱动的全流程自动化软件开发框架。ChatDev (Chat-powered Software Development)。ChatDev 拟作一个由多智能体协作运营的虚拟软件公司，在人类“用户”指定一个具体的任务需求后，不同角色的智能体将进行交互式协同，以生产一个完整软件（包括源代码、环境依赖说明书、用户手册等）。这一技术为软件开发自动化提供了新的可能性，支持快捷高效且经济实惠的软件制作，未来将有效地将部分人力从传统软件开发的繁重劳动中解放出来。

[GitHub - OpenBMB/ChatDev: Create Customized Software using Natural Language Idea \(through LLM-powered Multi-Agent Collaboration\)](#)



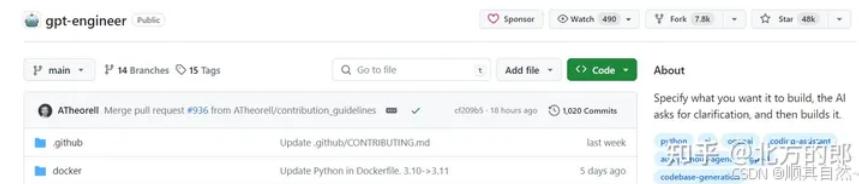
3.4、XAgent

GitHub - OpenBMB/XAgent: An Autonomous LLM Agent for Complex Task Solving

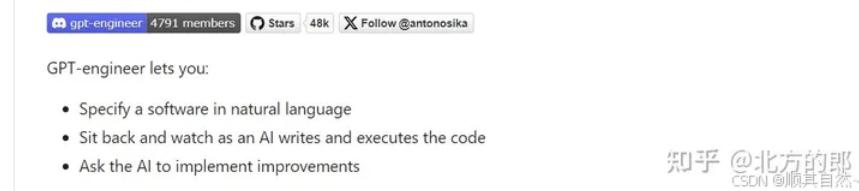


3.5、GPT-engineer

GitHub - gpt-engineer-org/gpt-engineer: Specify what you want it to build, the AI asks for clarification, and then builds it.



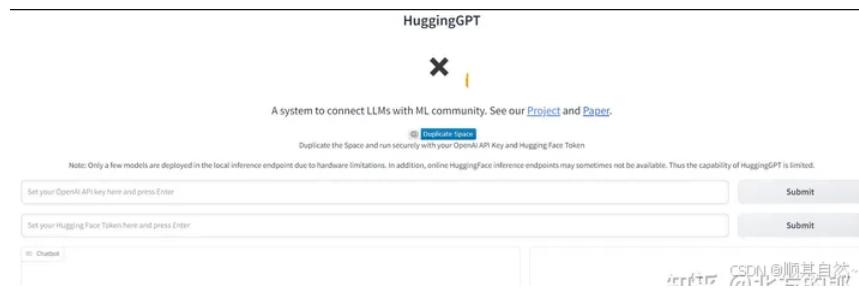
♂ GPT-Engineer



3.6、HuggingGPT

HuggingGPT也是一个老牌的AI Agent项目了，主要思路为利用LLM的框架（例如ChatGPT）来连接机器学习社区中的各种AI模型（例如huggingface）来解决人工智能任务。

网址：HuggingGPT - a Hugging Face Space by microsoft



代码：<https://github.com/>

JARVIS Public

main 4 Branches 0 Tags Go to file Add file Code About

tricktreat Merge pull request #228 from Wei-Zao/main · 93 Commits

huggingpt TaskBench 2 months ago

taskbench Add batch_evaluate.sh 2 days ago

.gitignore update 6 months ago

CITATION.cff update .cff 10 months ago

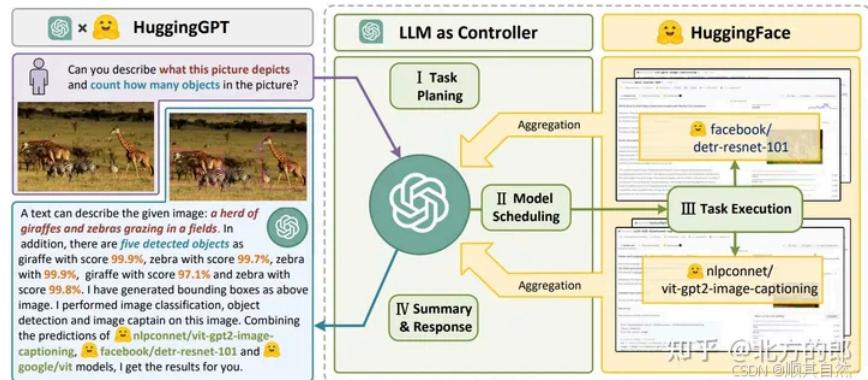
CODE_OF_CONDUCT.md CODE_OF_CONDUCT.md committed 10 months ago

CONTRIBUTING.md General Improvements to CONTRIBUTING.md (#55) 9 months ago

platforms deep-learning pytorch

Readme MIT license Code of conduct Security policy Activity

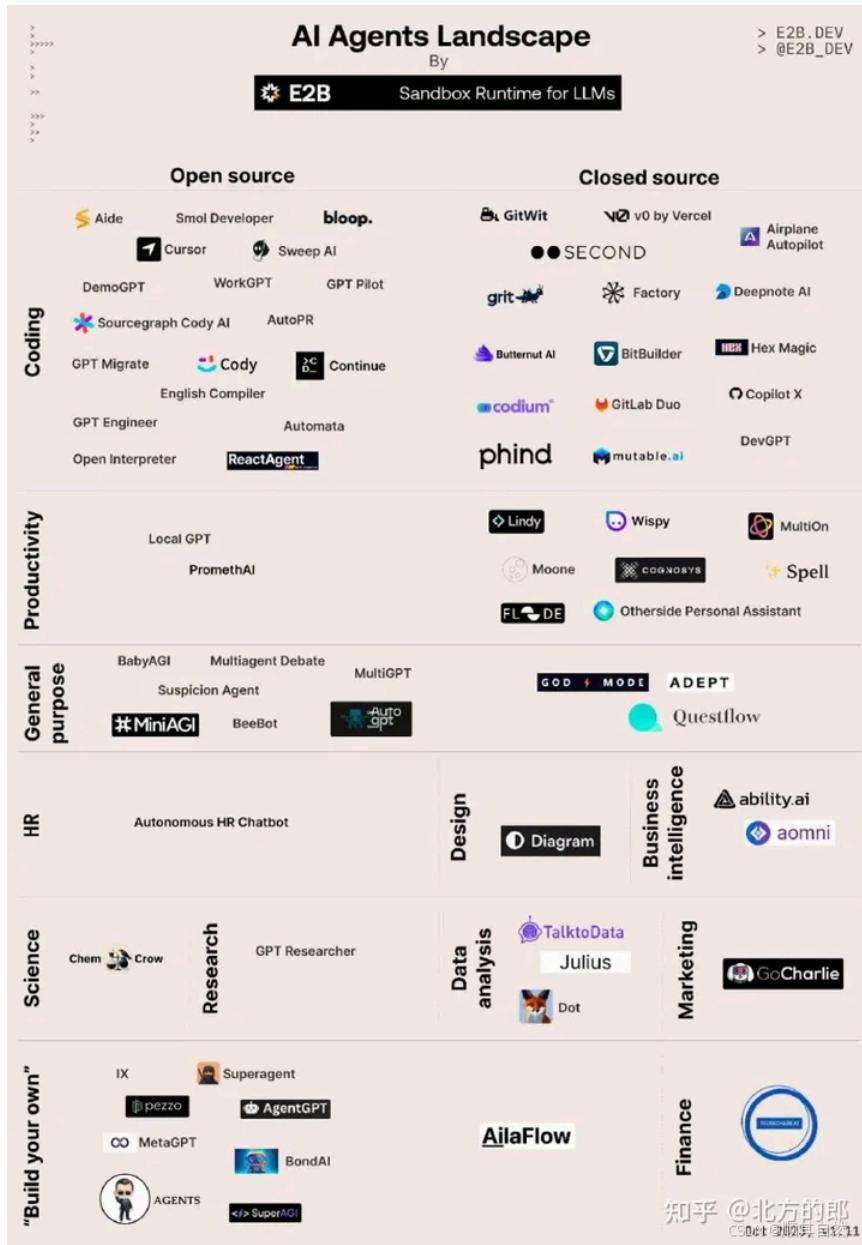
论文：HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face



4、AI Agent的展望与挑战

4.1、展望

AI Agent是人工智能成为基础设施的重要推动力。回顾技术发展史，技术的尽头是成为基础设施，比如电力成为像空气一样不易被人们察觉，但是又必不可少的基础设施，还如云计算等。当然这个要经历以下三个阶段：创新与发展阶段—新技术被发明并开始应用；普及与应用阶段—随着技术成熟，它开始被广泛应用于各个领域，对社会和经济产生深远影响；基础设施阶段—当技术变得普及到几乎无处不在，它就转变成了一种基础设施，已经成为人们日常生活中不可或缺的一部分。几乎所有的人都认同，人工智能会成为未来社会的基础设施。而智能体正在促使人工智能基础设施化。这不仅得益于低成本的Agent软件生产优势，而且因为Agent能够适应不同的任务和环境，并能够学习和优化其性能，使得它可以被应用于广泛的领域，进而成为各个行业和社会活动的基础支撑。



人工智能智能体应用一览图

Agent下一步可能会朝着两个方向同时迭代。一是与人协助的智能体，通过执行各种任务来协助人类，侧重工具属性；二是拟人化方向的迭代，能够自主决策，具有长期记忆，具备一定的类人格特征，侧重于类人或超人属性。

4.2、挑战

从技术优化迭代和实现上来看，AI Agent的发展也面临一些瓶颈。

技术方面，LLM模型仍然不够强大，即使是最强大的GPT4在AI Agent应用时，仍然面临以下一些问题：

上下文长度有限：上下文容量有限，限制了历史信息、详细说明、API调用上下文和响应的包含。系统的设计必须适应这种有限的通信带宽，而从过去的错误中学习的自我反思等机制将从长或无限的上下文窗口中受益匪浅。尽管向量存储和检索可以提供对更大知识库的访问，但它们的表示能力不如充分关注那么强大。

长期规划和任务分解的挑战：长期规划和有效探索解决方案空间仍然具有挑战性。LLM在遇到意外错误时很难调整计划，这使得它们与人类相比（从试错中学习）不太稳健。

自然语言接口的可靠性：当前的Agent系统依赖自然语言作为LLM与外部组件（例如内存和工具）之间的接口。然而，模型输出的可靠性值得怀疑，因为LLM可能会出现格式错误，并且偶尔会表现出叛逆行为（例如拒绝遵循指示）。因此，大部分Agent演示代码都专注于解析模型输出。

其次，太烧钱了，尤其是多智能体。斯坦福的虚拟小镇一个Agent一天需要消耗20美金价格的token数，因为其需要记忆和行动的思考量非常大。这一价格是比很多人类工作者更高的，需要后续Agent框架和LLM推理侧的双重优化。

还有就是现阶段在很多场景，使用AI Agent对比Copilot模式的最终效果，还看不到非常大的提升，或者说能覆盖增加成本的提升。大部分AI Agent技术还都是研究阶段。

最后，这些发展趋势预示着AI Agent可能面临诸如安全性与隐私性、伦理与责任、经济和社会就业影响等多方面的挑战。别的不说，对很多人的个人职业生涯的长期影响。

Generative AI is already taking white collar jobs and wages in the online freelancing world

Change in employment and earnings from writing and editing jobs on an online freelancing platform after the launch of ChatGPT



Source: *The Short-Term Effects of Generative AI on Employment: Evidence from an Online Labor Market* (Hu et al., 2023)
© FT

知乎 @北方的郎
CSDN @顺其自然~

以ChatGPT的发布为分水岭，全球自由职业平台上的写作/编辑类从业者的数量和收入都进入了断崖式下跌的轨道

参考文献：

<https://lilianweng.github.io/posts/2023-06-23-agent/>

<https://arxiv.org/abs/2312.17025>

<https://developer.nvidia.com/blog/introduction-to-lm-agents/>

<https://www.gatesnotes.com/AI-agents>

AI Agent，为什么是AIGC最后的杀手锏？

<https://github.com/e2b-dev/awes>

转自：<https://fuhanghang.blog.csdn.net/?type=blog>