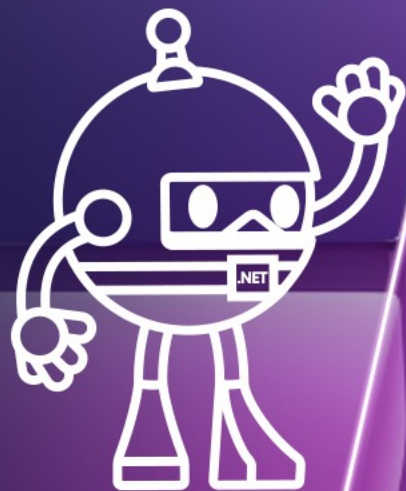


.NET Conf China 2023

2023/12/16
09:30 - 18:00

中国 · 北京



中国·北京

.NET Conf China 2023

多模型加持下的
.NET Copilot 架构

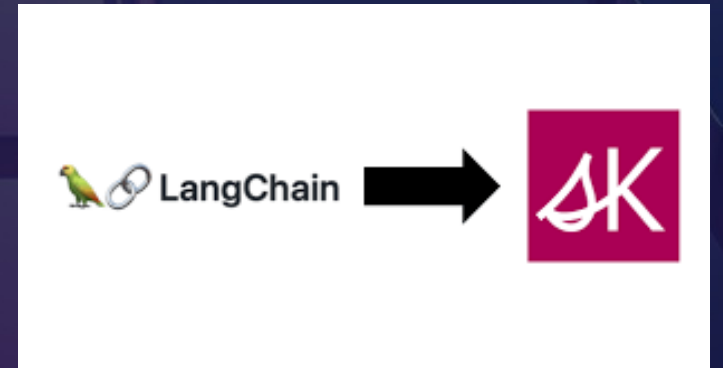




Prompt 工程很重要

很多新的论文发表

应用框架



入坑后，各种水土不服



什么是智能化转型



聊天机器人



Copilot



AI Agent



大模型的特点

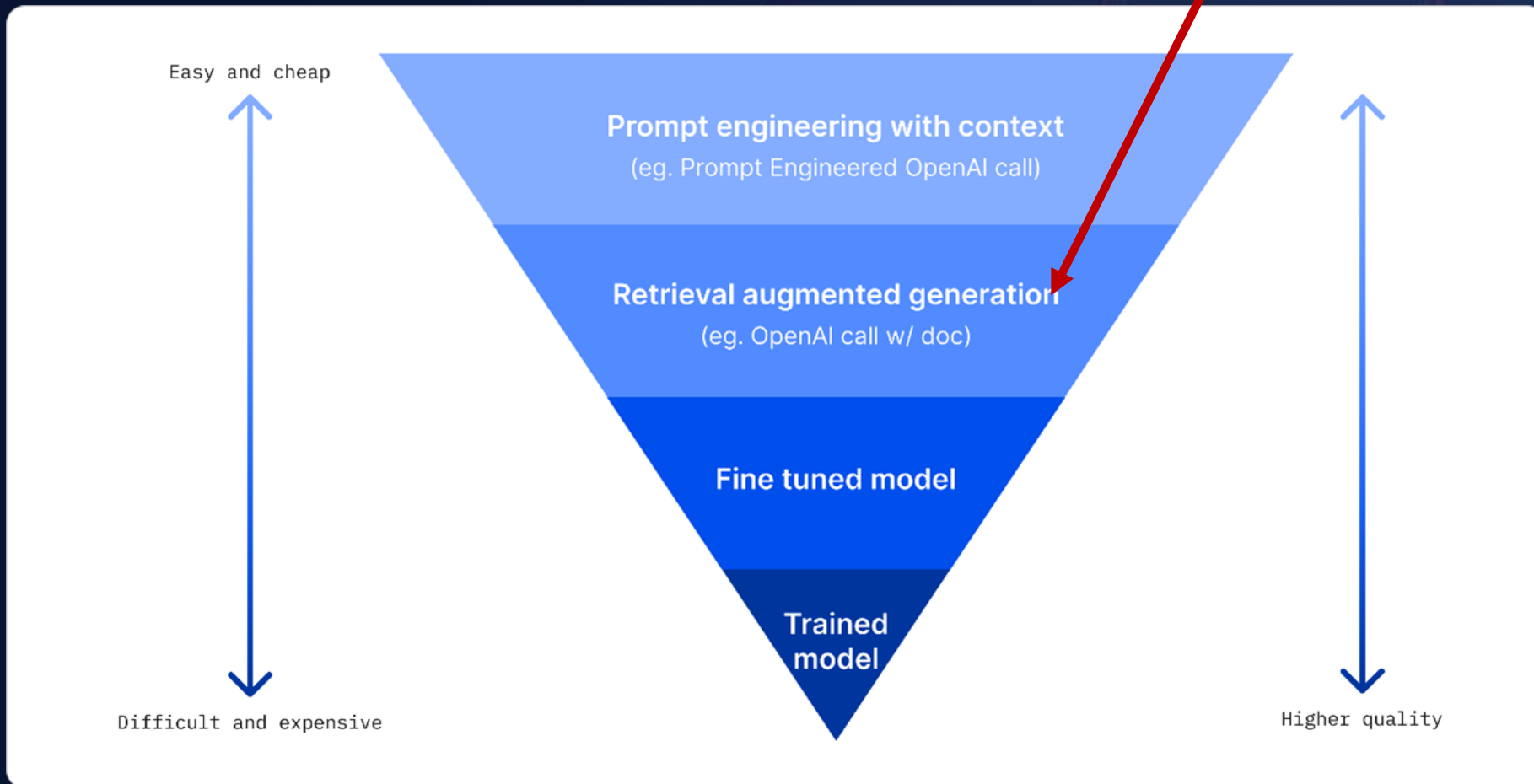
梦想

幻觉

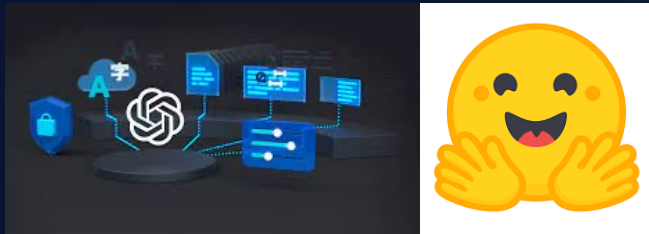


可以解决 80% 的问题

实际落地的故事



但实际情况，你可以看到



API or Model



Data & KB



Copilots



模型选择对企业来说是两难



多模型的生态是必不可少的 特别在国内的环境下



如何接入?

完全 .NET 的解决方案

LLaMa Sharp

SciSharp/LLamaSharp

Local Hugging Face Model Inference Server 方案

Local LLM Server for Semantic Kernel .NET Developer

如果你使用中文，请访问[这里](#)

Support Semantic-Kernel RC.3

Or you are using Semantic Kernel's [Hugging Face http server](#) as your local LLM service , but based on the inaccessibility of hugging face in mainland China and management reasons, I tried to reconstruct the project. At this stage, it is adapted for macOS and Linux environments.

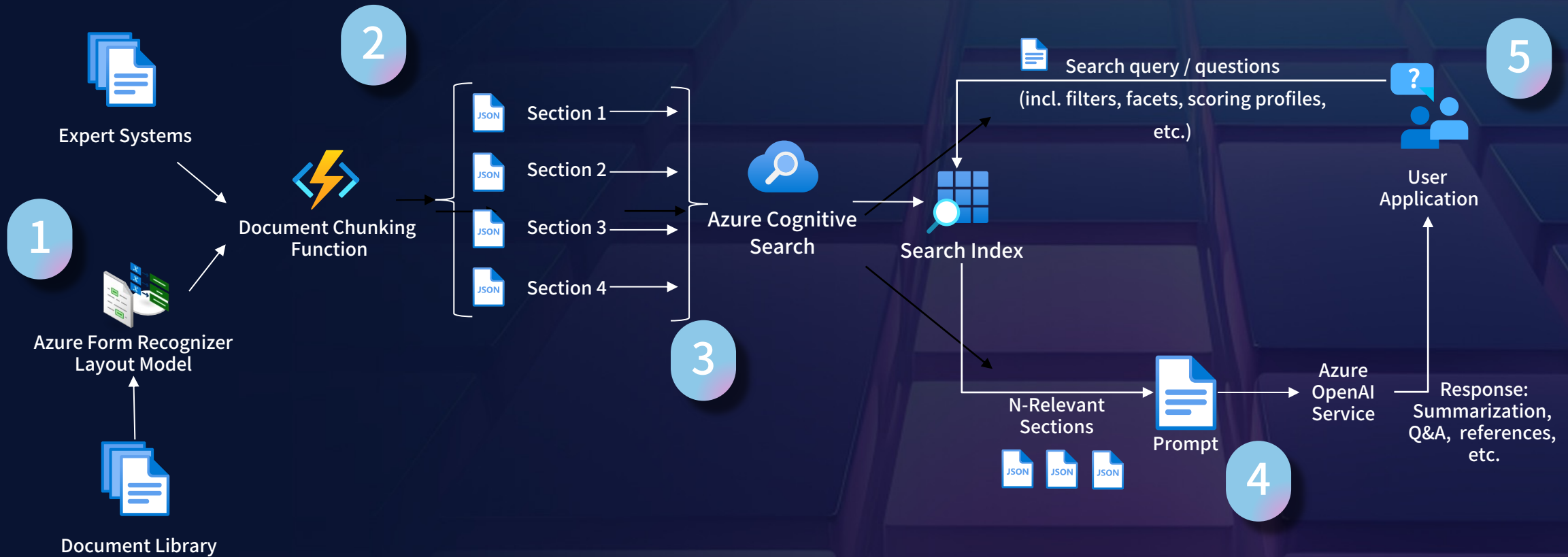
At this stage, the implementation of ChatCompletion and Embedding has been completed.

ChatCompletion is adapted to LLM *baichu2*, *ChatGLM3*, *Microsoft-phi 1.5*

Embeddings adapted to LLM *jina-embeddings* (English Embedding model), *text2vec-large-chinese* (Chiense Embedding model), *bge-*-v1.5*(Chinese/English Embedding model)

[kinfey/SemanticKernel-Local-LLM-Server](#)





1. 数据提取

不同的数据格式

2. 分块

最好的分块策略是什么？

3. 索引

应该使用向量、语义还是传统方法？

4. Prompting

工具、技巧和策略

5. 用户界面

如何展现信息？



离线大模型能做吗？

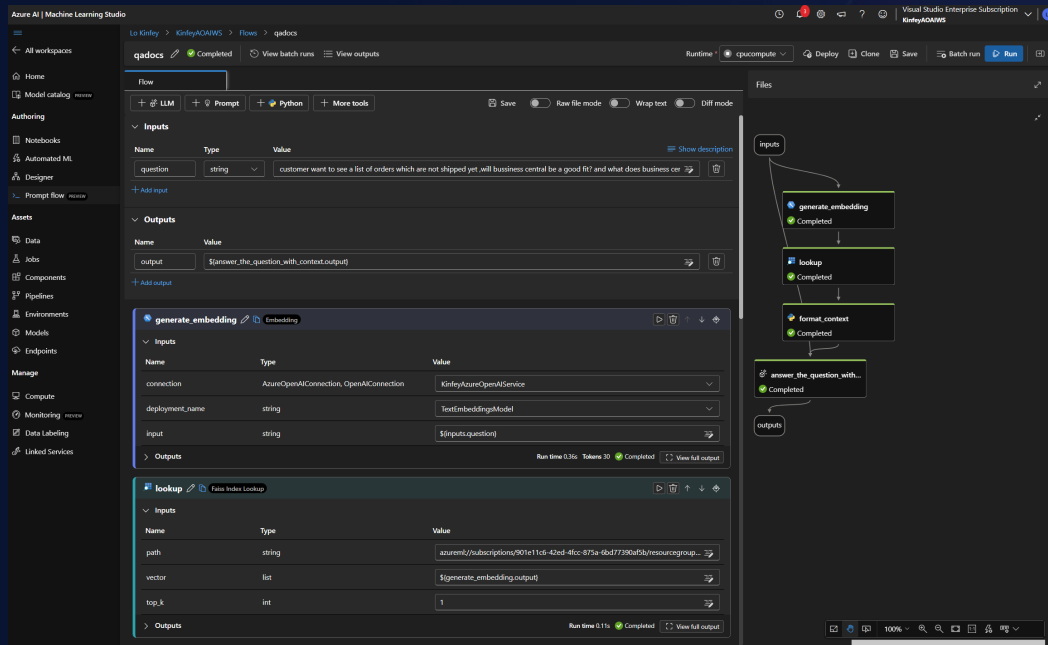
Demo



幻觉解决的方法是否有？



- ✓ 平台
- ✓ 针对调整和迭代 LLM 支持的工作流程进行了优化



Azure Machine Learning Prompt Flow

Azure Machine Learning Prompt Flow 是一款开发工具，旨在简化由大型语言模型 (LLM) 提供支持的 AI 应用程序的整个开发周期，提供全面的解决方案来简化 AI 应用程序的原型设计、实验、迭代和部署过程。

Promptflow 做了啥

- 增强开发工作流程
 - 开发连接到各种语言模型、外部数据源、工具和自定义代码的流程
- 测试和评估
 - 使用大型数据集并行测试流程
 - 使用性能、基础性和准确性等指标评估工作流程的 AI 质量
- 提示工程调整
 - 轻松调整提示工程和版本
- 比较和部署
 - 比较不同的结果
 - 一键部署到托管端点以实现快速集成



对于固定业务流就是简单固定的 Prompt



.NET 版本的 Prompt flow 即将到来

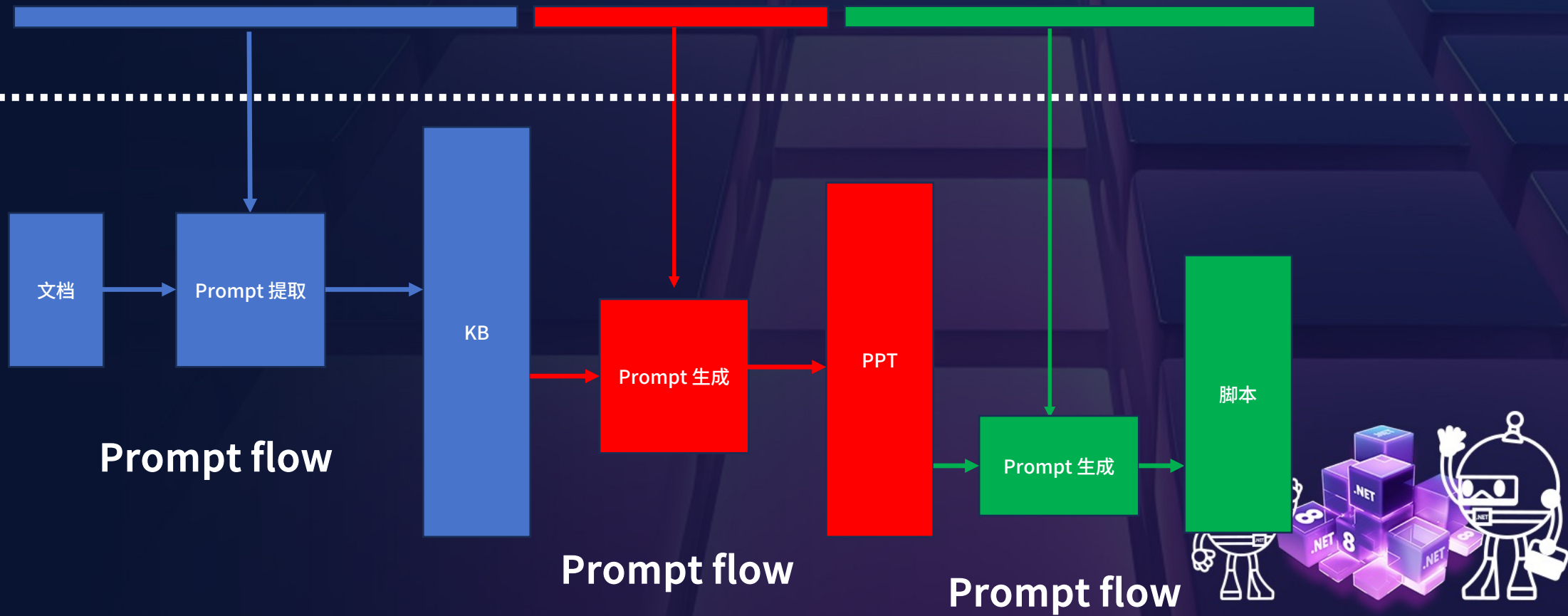




有了 Prompt flow 后的应用更进一步

前端 帮我根据书本的内容，生成 PPT，并生成每一页的脚本

后端





Enterprise Copilot @Copilot Stack



VSCoDe Extensions

Copilot frontend

Plugin extensibility

UX

Skills



- Enterprise Code Analyze
- Enterprise Code Testing
- Enterprise Code Testing
- Enterprise Code Generation
- Enterprise Docs Generation
- Enterprise Arch Generation



PromptFlow

Orchestration

Prompt & response filtering

Metaprompt

Grounding

Plugin execution



Azure OpenAI Services



Code-Llama

Foundation models

Hosted foundation models

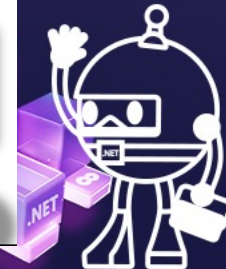
Hosted fine-tuned foundation models

BYO models



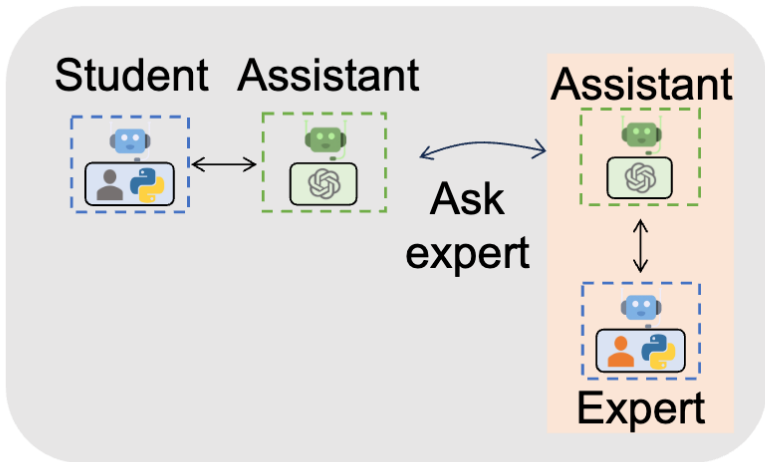
Azure Machine Learning

AI infrastructure

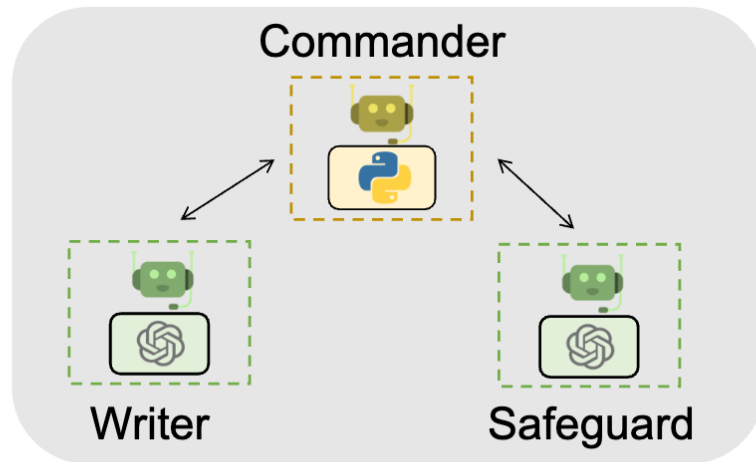


AI Agents 让智能再进化，综合我们的框架进行思考

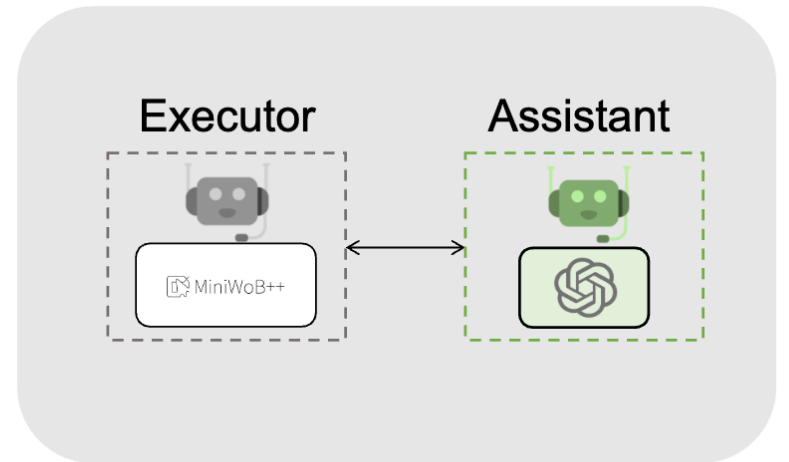




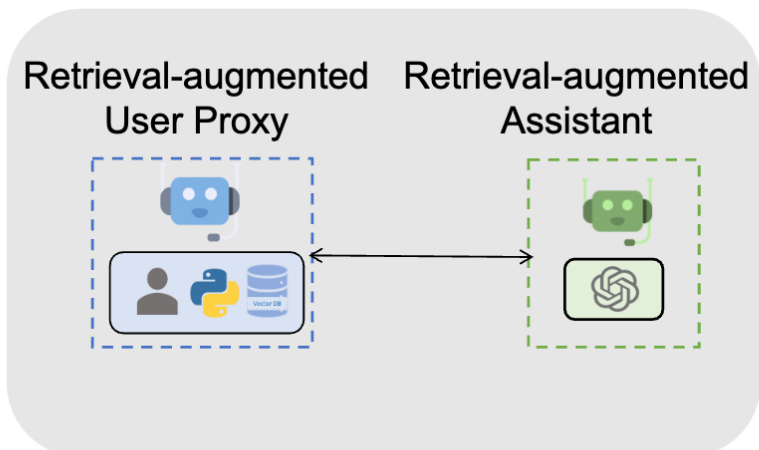
A1. Math Problem Solving



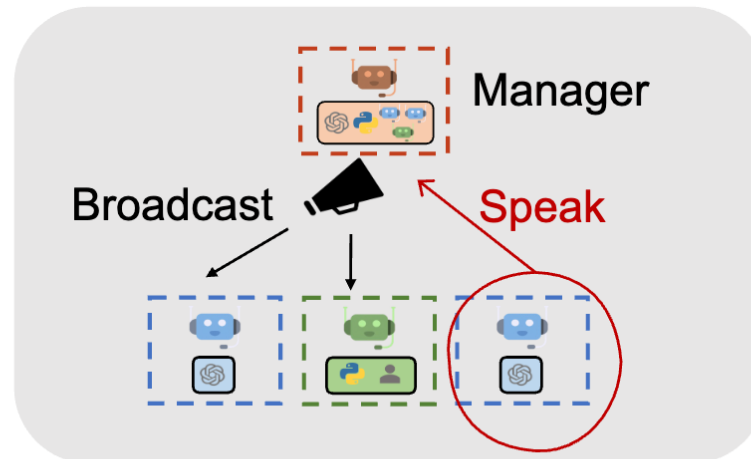
A2. Multi-agent Coding



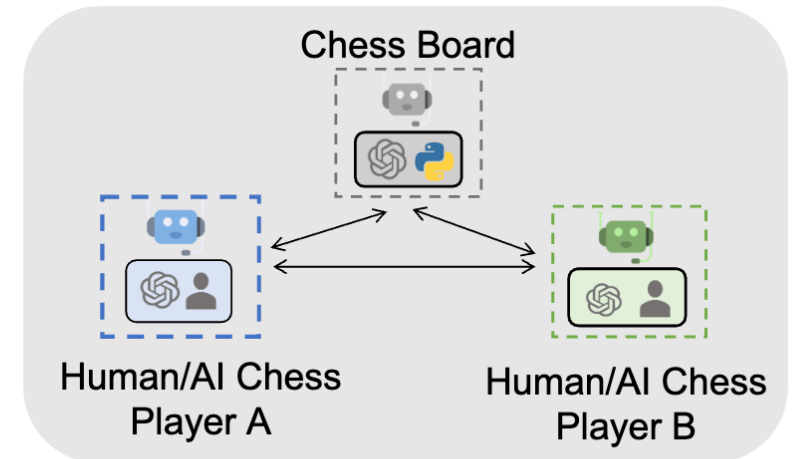
A3. Online Decision Making



A4. Retrieval-augmented Chat



A5. Dynamic Group Chat



A6. Conversational Chess



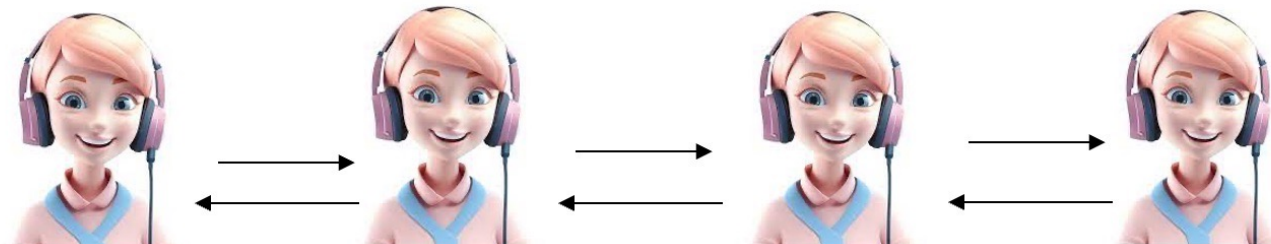


HuggingGPT



AI Agents 派工

目标：根据需求文档帮我生成一个 iOS 的移动应用



需求文档 agent

代码生成 agent

测试代码 agent

发布应用 agent

读取文档，做出归纳，把每一步的内容列出

创建项目，根据每一步生成对应的代码

运行项目，生成不同的测试用例

测试通过后，打包应用，并进行签名发布到企业商店



Semantic Kernel / LangChain 任务拆解



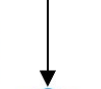
阅读技能



归纳技能



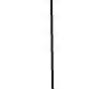
拆解技能



iOS 技能



代码生成技能



运行技能



测试生成技能



自动测试技能



应用打包技能



签名技能



发布技能

Prompt flow 解决的是 Prompt 和大模型交互后输出稳定性的问题

多模型结合的生态



Thanks

