

# 大语言模型在经济学中的应用

## 从 API 调用出发的学术应用实例

陆知雨

中国财政协同发展研究中心

财政学基地班 22

2025 年 3 月 9 日

引言

基于代码层面  
的 LLM 调用

如何强化 LLM  
在特定领域的  
理解力

在经济学领域  
应用的具体案  
例

基于 LLM 的  
开源项目分享

- ① 引言
- ② 基于代码层面的 LLM 调用
- ③ 如何强化 LLM 在特定领域的理解力
- ④ 在经济学领域应用的具体案例
- ⑤ 基于 LLM 的开源项目分享

引言

基于代码层面的  
LLM 调用

如何强化 LLM  
在特定领域的  
理解力

在经济学领域  
应用的具体案  
例

基于 LLM 的  
开源项目分享

## 1 引言

## 2 基于代码层面的 LLM 调用

## 3 如何强化 LLM 在特定领域的理解力

Prompt Engineering（提示词工程）

RAG（检索增强生成）

Fine-Tuning（微调）

## 4 在经济学领域应用的具体案例

## 5 基于 LLM 的开源项目分享

Chatbox & CherryStudio

PDF2zh

Ollama

其他

# 背景与目标

- 什么是大语言模型？大语言模型（LLM）的本质在于：在大量向量化的词语库的基础上，通过数学模型实现词语预测的功能，并且这一数学模型可以被抽象为 LLM 对于语言含义的“理解”。
- 对于经济学领域的影响？文本数据是经济学研究中最常见的数据类型之一，LLM 基于文本训练的特性可以使其以更低成本实现对文本数据的处理。
- 本次分享的安排如下：
  - ① 基于代码层面调用大语言模型处理文本数据的方法与学术实例
  - ② 使用 RAG 与 Finetuning 特化 LLM 在某一领域理解力的方法与实例
  - ③ 基于大语言模型的实用开源项目介绍

## ① 引言

## ② 基于代码层面的 LLM 调用

## ③ 如何强化 LLM 在特定领域的理解力

Prompt Engineering（提示词工程）

RAG（检索增强生成）

Fine-Tuning（微调）

## ④ 在经济学领域应用的具体案例

## ⑤ 基于 LLM 的开源项目分享

Chatbox & CherryStudio

PDF2zh

Ollama

其他

# 为什么使用 API 调用 LLM

- API 指应用程序编程接口，通过 API 用户可以网络链接提供商的服务器，从而在代码中嵌入 LLM 功能。

API 调用 LLM	网页端使用 LLM
代码交互	可视化交互
需要有一定编程基础	容易上手
应用场景灵活	限于固定功能与场景
支持对模型进行调整	仅提供基础选择
支持长文本批量处理	仅支持单文本
按使用量计费	免费或订阅制付费

# 主流大语言模型 API 的获取

## 引言

### 基于代码层面的 LLM 调用

### 如何强化 LLM 在特定领域的 理解力

### 在经济学领域 应用的具体案 例

### 基于 LLM 的 开源项目分享

- **国内厂商:** 直接登陆官方网站注册账号即可获取。
- **国外厂商:** 通常需要通过境外信用卡支付, 对网络条件有一定要求。

解决方案 (仅作参考):

- ① 使用虚拟信用卡<https://yeka.ai>, 优化网络环境。
- ② 使用国内代理商服务商<https://o3.fan>

## API 参考定价 (2025-03-09)

## 引言

基于代码层面的  
LLM 调用

如何强化 LLM  
在特定领域的  
理解力

在经济学领域  
应用的具体案  
例

基于 LLM 的  
开源项目分享

机构	模型	输入/输出价格 (美元/百万 tokens)
阿里	Qwen-max	1.6 / 6.4
	Qwen-plus	0.4 / 1.2
OpenAI	GPT-4o	2.5 / 10
	GPT-4o-mini	0.15 / 0.6
	o1	15 / 60
谷歌	Gemini 2.0 Flash	0.1 / 0.4
	Gemini 2.0 Flash-Lite	0.075 / 0.3
	Gemini 1.5 Flash	0.075 / 0.3
Anthropic	Claude 3.7 Sonnet	3 / 15
	Claude 3.5 Haiku	0.8 / 4
	Claude 3 Sonnet	15 / 75
深度求索	deepseek-chat	0.27 / 1.1
	deepseek-reasoner	0.55 / 2.19
月之暗面	kimi-latest (8k-128k)	1.62-8.1 / 1.62-8.1
字节跳动	Doubao-pro (4k-256k)	0.1-0.675 / 0.27-1.215

注意：实际使用中，不同厂商有不同优惠政策，成本可能更低。

输入价格：上传至模型的数据收费；输出价格：模型返回结果的数据收费。

**tokens** 为模型处理文本的基本计量单位，通常一个汉字约等于 0.6 个 tokens，一个英文单词约等于 0.3 个 tokens。



# API 调用的代码示例 (Python)

使用 **openai** 库: 目前主流厂商的 API 调用都可以通过 **openai** 库实现, 因此仅展示 **openai** 库的使用方法。

---

```
from openai import OpenAI
client = OpenAI() # 初始化OpenAI客户端, 可填入API密钥与url网址
prompt = "You are a helpful assistant" # 系统角色提示
input_text = "What is the capital of France?" # 用户输入问题
completion = client.chat.completions.create(
    model="gpt-4o", # 指定模型
    messages=[
        {"role": "system", "content": prompt}, # 系统角色消息
        {"role": "user", "content": input_text} # 用户角色消息
    ],
    temperature=1, # 温度1为最高
    max_tokens=2048, # 限制生成最大长度
    top_p=1, # 采样策略1为全概率
    frequency_penalty=0, # 禁用重复惩罚
    presence_penalty=0 # 禁用主题惩罚
)
print(completion.choices[0].message.content) # 输出首个响应内容
#输出示例: The capital of France is Paris.
```

---

# API 调用进阶 1: 重要参数解析

## 引言

基于代码层面的  
LLM 调用

如何强化 LLM  
在特定领域的  
理解力

在经济学领域  
应用的具体案  
例

基于 LLM 的  
开源项目分享

- **model**: 指定模型版本, 如"gpt-4o"、"deepseek-chat"。
- **messages**: 构建对话历史 (系统、用户、助手)。
- **temperature**: 控制随机性, 数值越高生成越多样。
- **max\_tokens**: 限制返回文本的最大长度。
- **top\_p**: 核采样策略, 1 为全概率采样。
- **frequency\_penalty** & **presence\_penalty**: 调整重复率和新话题引入。

**建议:** 根据任务需求调整参数, 如创意写作可提高 temperature, 事实查询则建议降低 temperature。具体设置, 可参考

## API 调用进阶 2: 多线程调用

### 引言

基于代码层面的  
LLM 调用

如何强化 LLM  
在特定领域的  
理解力

在经济学领域  
应用的具体案  
例

基于 LLM 的  
开源项目分享

在实际应用中，为了提高 API 调用效率，可以采用多线程并发处理多个请求。多线程调用的优势包括：

- **提高响应速度：**多个线程可以同时发起请求，充分利用系统资源，从而减少总体等待时间。
- **更高的吞吐量：**并发执行多个 API 调用可以显著提升整体处理能力，适用于高并发场景。
- **简化代码结构：**利用现有的多线程库（如 `concurrent.futures`）可以方便地管理线程，减少手动管理线程的复杂性。

具体代码示例，已保存在项目库中，可进一步查看。

**注意：**由于大部分平台都对访问流量设置限制，使用多线程时需要特别注意线程安全和异常处理。

## ① 引言

## ② 基于代码层面的 LLM 调用

## ③ 如何强化 LLM 在特定领域的理解力

Prompt Engineering (提示词工程)

RAG (检索增强生成)

Fine-Tuning (微调)

## ④ 在经济学领域应用的具体案例

## ⑤ 基于 LLM 的开源项目分享

Chatbox & CherryStudio

PDF2zh

Ollama

其他

# Prompt 工程

- 为什么还需要 **prompt**? 虽然目前最先进大语言模型已内置思维链, 使用时无需复杂提示, 但经济学领域的任务多为文本分类和语义识别。此时, 采用思维链模型成本较高且优势不明显, 因此应选择效果好且经济的方案。
- 什么是好的 **prompt**? 好的 prompt 应具备以下特点:
  - ① 明确任务目标。
  - ② 提供清晰的输入。
  - ③ 明确期望的输出格式。
  - ④ 提供思维链。
  - ⑤ 指令保证格式清晰

## 小贴士:

- 关于各类具体情境下的 prompt 示例, 可参考<https://github.com/openai/openai-cookbook/>
- 众多 API 平台都提供了 prompt 生成功能, 可根据需求进行选择。

# 构建知识库

**为什么 RAG:** 目前主流 API 平台，如：阿里、coze，大多支持上传 pdf、word、txt 等格式的文档，通过将文本向量化构建知识库，再通过检索与生成的方式实现对特定领域的文本理解。

与直接使用大语言模型相比，RAG 具有以下优势：

- ① **成本更低：**相较于直接调用大语言模型，RAG 只需要在初始阶段对文档进行一次性向量化，后续生成过程仅依赖于检索结果，从而降低长期使用的成本。
- ② **更高效：**利用文档中的信息作为上下文，能提高生成回答的准确性和完整性，避免因缺乏领域知识而产生的错误或偏差。
- ③ **更灵活：**RAG 可以根据具体需求调整检索和生成模块，同时适应不断更新的领域知识，满足实时性要求。

# 知识库 vs 非知识库回答

问题：简要概括什么是企业绿色发展？

知识库：由 10 篇关于企业绿色发展的论文组成。

	使用知识库	不使用知识库
回 答	企业绿色发展是指企业以绿色发展理念为指导，通过资源集约利用和环境友好为导向，以绿色创新为核心，实现生产全过程的绿色化，同时兼顾经济绩效和环境绩效，最终达到生态环境改善和经济社会高质量发展的目标。这一过程涉及环境绩效、经济绩效和绿色创新绩效三个方面。环境绩效表现为企业在绿色转型过程中为保护环境和治理污染所取得的成绩；经济绩效是企业通过资源分配与利用所获得的效益；绿色创新绩效则是企业通过绿色创新积蓄的技术驱动力。	企业绿色发展是指企业在生产经营过程中，通过采用环保技术、优化资源利用、减少污染排放、降低能源消耗等措施，实现经济效益与环境保护的双赢。其核心在于将可持续发展理念融入企业战略和日常运营中，推动经济增长与生态环境保护的协调统一。企业绿色发展不仅有助于减少对环境的负面影响，还能提升企业竞争力，满足社会对环保的期望，促进长期可持续发展。

分析：同一个模型，引入知识库后，回答更倾向于目前学术研究的关注方向，与不使用知识库的回答有较大区别。

# 通过微调强化特定领域能力

## 引言

基于代码层面的  
LLM 调用

如何强化 LLM  
在特定领域的  
理解力

Prompt Engineering  
(提示词工程)

RAG (检索增强生成)

Fine-Tuning (微调)

在经济学领域  
应用的具体案  
例

基于 LLM 的  
开源项目分享

**为什么微调?** 微调模型是目前经济学论文中常用的步骤，通过微调可以强化模型在特定领域的表现，确保分析结果的可信度和针对性。

**微调步骤：**

- ① 准备并上传训练数据
- ② 训练新的微调模型
- ③ 评估结果并根据需要返回步骤 1
- ④ 使用微调模型



## 如何准备微调数据:

- ① **数据标注**: 对于需要监督学习的任务, 需要对数据进行准确的标注:
- ② **数据格式化**: 将预处理后的数据转换为模型所需的格式。常见格式有 JSON、CSV 等。例如, 在基于文本分类的微调中, 由于要求回答相对单一, 可以采用如下格式:  

```
{"prompt": "请你判断采购大数据云监控平台数据//  
属于集权还是分权类采购?", "completion": "集权"}
```
- ③ **数据扩增 (可选)**: 当可用数据量不足时, 可考虑使用数据扩增技术, 如同义句生成、数据混合或噪声注入等方法, 增加数据多样性, 提高模型鲁棒性。

关于微调的具体介绍, 已保存在项目库中, 可进一步查看。

# 微调测试案例

## 测试案例一：

采购公告：怒江州林业局森林防火地理信息指挥系统采购项目中标公示

是否微调	是否需要 prompt	模型名称	分类结果
是	否	ERNIE 4.0	两者
否	是	ERNIE 4.0	(未能识别任务)
否	是	GPT-4o	两者
否	是	GPT-4o-mini	两者

## 结果发现：

微调后的 ERNIE 4.0 在识别准确性上大致接近于未经过微调的 GPT-4o-mini。

后续查询相关论坛帖子，发现如果只是对文本进行分类处理，只要选择足够优秀的大语言模型，微调并不必要。

## ① 引言

## ② 基于代码层面的 LLM 调用

## ③ 如何强化 LLM 在特定领域的理解力

- Prompt Engineering（提示词工程）
- RAG（检索增强生成）
- Fine-Tuning（微调）

## ④ 在经济学领域应用的具体案例

## ⑤ 基于 LLM 的开源项目分享

- Chatbox & CherryStudio
- PDF2zh
- Ollama
- 其他

# 案例一：大语言模型、文本情绪与金融市场（姜富伟等，2024）

## 主要研究内容：

- 构建中文金融大语言模型，用于预测金融市场情绪、资产价格回报等。
- 将结构化的市场数据（如股价、回报率）与非结构化文本数据（如财经新闻、公司公告）相结合，通过大模型更精准地刻画金融市场非理性情绪。

## 核心技术方案：

- 以 BERT/ERNIE 类深度学习模型为基础做预训练与微调。
- 引入市场回报标签，进行有监督的情绪分类或回归预测。
- 对比传统字典法，验证“大模型情绪”在极端市场状况下的预测优越性。

## 案例一：模型微调（Fine-tuning）

- 思路：
  - 在通用预训练模型基础上，利用市场回报数据作为情绪标签进行有监督微调。
- 实现过程：
  - ① 选定预训练模型：如 BERT、ERNIE、RoBERTa 等，下载权重。
  - ② 标注数据集：以股票收益正负为情绪标签（0/1 分类），或多档位标签。
  - ③ 训练设置：设定 batch size、learning rate、epoch 等参数，使用 GPU 训练。
  - ④ 模型评估：分类准确率、F1-score、在样本外的收益预测表现等。
  - ⑤ 模型部署：将微调好的模型用于未来新闻文本情绪批量处理。
- 效果：
  - 明显优于传统字典法，捕捉到非线性、上下文关联和极端事件下情绪冲击。

## 案例二：企业数字化转型的测度难题 (金星烨等, 2024)

### 主要研究内容：

- 提出用大语言模型识别上市公司年报中是否真实应用数字技术（如 AI、大数据、云计算等），从而构建更准确的数字化转型指标。
- 解决传统词典法关键词遗漏和误判等问题，进而研究数字化转型对企业绩效的影响差异。

### 核心技术方案：

- 采用大语言模型（如 ERNIE）对年报句子逐句判定是否涉及数字技术。
- 与人工标注结合，“进一步预训练” + 微调，在构造新指标时更加完备且准确度更高。

## 案例二：模型微调（Fine-tuning）

- 数据标注：
  - 提前选定一个人工标注数据集（部分年报句子），区分是否真实采用数字技术。
- 微调过程：
  - ① 在已有的中文模型（如 ERNIE）上进行“进一步预训练”（further pretraining），让模型适应企业年报用语。
  - ② 基于标注数据进行分类微调：句子 → 是否包含“真实使用数字技术”。
  - ③ 最终模型在测试集上评估指标，如 Accuracy、Precision、Recall、F1-score 等。
- 效果：
  - 相较于关键词检索法，显著降低误判（谈技术却未实际应用）与漏判（实际应用技术却关键词缺失）的概率。

## ① 引言

## ② 基于代码层面的 LLM 调用

## ③ 如何强化 LLM 在特定领域的理解力

Prompt Engineering（提示词工程）

RAG（检索增强生成）

Fine-Tuning（微调）

## ④ 在经济学领域应用的具体案例

## ⑤ 基于 LLM 的开源项目分享

Chatbox & CherryStudio

PDF2zh

Ollama

其他



# Chatbox & CherryStudio：通过 API 实现各平台 LLM 集成

- **概述：**基于 API 调用机制，可以通过上传 API 的形式，实现各平台 LLM 的集成。
- **优势：**将 API 调用可视化实现类似于网页的功能；热门模型如 deepseek 网页端经常卡顿，但是使用 API 调用则不会。

# PDF2zh: 将 LLM 引入 pdf 文献 翻译

- 已有翻译引擎仅限于机翻，少部分可以使用 AI 的翻译软件并不能提供全文翻译
- 此项目可以讲 PDF 全文翻译为中文，并保留排版格式，方便阅读。
- 由于项目自身基于命令行，且其可视化窗口配置较为复杂。基于此项目开发了可视化项目，也可供参考：

<https://github.com/luzhiyu-econ/PDF-Translator>

# Ollama: 本地开源大模型部署

## 引言

基于代码层面的  
LLM 调用

如何强化 LLM  
在特定领域的  
理解力

在经济学领域  
应用的具体案  
例

基于 LLM 的  
开源项目分享

Chatbox &  
CherryStudio

PDF2zh

Ollama

其他

- 依据本地计算机算力，主要受电脑显卡显存于电脑内存限制，可以在本地部署不同的 LLM 模型。并开放 API 供用户调用。
- 通过本地调用模型一方面节省了 API 费用；另一方面，也规避了后续官方模型调整，原实验结果不可复现的局面。

## 引言

基于代码层面的  
LLM 调用

如何强化 LLM  
在特定领域的  
理解力

在经济学领域  
应用的具体案  
例

基于 LLM 的  
开源项目分享

Chatbox &  
CherryStudio

PDF2zh

Ollama

其他

- Cursor, Trae: AI powered IDE
- 将 AI 功能集成到 IDE 中，可以实现代码生成、代码解释、代码调试等功能。
- Storm: 个人知识库构建
- 类似于平台提供的知识库功能，通过提供的文献资料，通过调用 LLM 接口，生成结构化的报告或新内容。

# References



Dell, M. (2023).  
Deep Learning for Economists.  
*Journal of Economic Literature*.



Korinek, A. (2023).  
Generative AI for Economic Research: Use Cases and Implications for  
Economists.  
*Journal of Economic Literature*.



Hong, Y. & Wang, S. (2024).  
ChatGPT 与大模型将对经济学研究范式产生什么影响?  
计量经济学报.