自动 Related Paper Agent 实现方案

# 1. 整体目标

做一个自动 related paper agent，输入主题/种子论文 → 自动检索 → 本地语义排序 → GPT 总结理由 → 输出报告。

# 2. 架构设计

用户输入 (课题 / 种子论文)  
 │  
 ▼  
[Query 生成与优化] —— GPT  
 │  
 ▼  
[开放数据源检索] —— OpenAlex / Semantic Scholar / arXiv API  
 │  
 ▼  
[语义嵌入召回] —— 本地 E5 / BGE 模型 + Faiss 向量索引  
 │  
 ▼  
[重排 Rerank] —— 本地 BGE-reranker  
 │  
 ▼  
[结果解释与总结] —— GPT (自动生成“相关理由”)  
 │  
 ▼  
输出 (Markdown/表格/PDF)

# 3. 模块拆分与实现

## 3.1 输入 & Query 优化

输入：用户输入课题主题或种子论文；用 GPT 生成英文关键词、同义词扩展，输出多个检索式，保证召回率。

## 3.2 数据源检索

调用 OpenAlex、Semantic Scholar、arXiv API 获取论文元数据。Python requests 实现，结果存 JSON/SQLite。

## 3.3 向量召回

用 HuggingFace E5/BGE 嵌入模型生成向量，Faiss 建立索引，召回候选文献。

## 3.4 重排

用本地 BGE-reranker 对候选论文与 query 做交互式重排，选出 Top-20。

## 3.5 结果解释与总结

调用 GPT 自动生成 Why Related / Difference，输出 Markdown 表格。

## 3.6 输出与展示

生成 Markdown 表格，或导出 PDF/Word。

# 4. 技术栈

- Python （主语言，Codex 辅助写代码）  
- Requests （API 调用）  
- Transformers （嵌入模型 + Reranker）  
- Faiss （向量索引）  
- SQLite / Pandas （存储与管理）  
- GPT Plus （Query 重写 + 结果解释）

# 5. 扩展功能

- 增量更新：自动每周抓取新文献  
- 引用图可视化：NetworkX + Matplotlib  
- Paper Summary 卡片：每篇论文自动生成三句话摘要

# 6.展示效果

1. 架构图（输入 → 检索 → 召回 → 重排 → GPT → 输出）  
2. Demo 截图（用户输入课题 → 输出表格）  
3. 对比传统检索与 Agent 的优势  
4. 强调“开源 + 本地运行 + GPT Plus 辅助解释”，无额外成本