

# HR 产假智能助手 🤖

专为HR产假管理设计的AI智能助手

专门为人力资源部门产假政策管理和员工咨询而设计的智能助手。通过自然语言对话，轻松处理产假政策文档、员工申请数据分析，以及各地区产假规定的查询比较，让HR工作更高效、更准确。

## ✨ HR产假智能助手能做什么？

### 📄 产假政策文档智能分析

- 政策文档对话：上传各地产假政策文件，用中文直接提问政策细节
- 多格式支持：支持PDF、Word、Excel等格式的政策文档和员工数据
- 智能政策搜索：快速查找特定地区、特定情况的产假规定
- 批量政策比较：同时分析多个地区的产假政策差异

### 📊 产假数据分析与统计

- 员工数据查询："哪些员工即将休产假？"或"本季度产假申请趋势如何？"
- 政策对比分析："北京和上海的产假天数有什么区别？"
- 数据统计报告：自动生成产假申请统计、费用预算等报告
- 合规性检查：验证员工产假申请是否符合当地政策规定

### 🧠 专业HR智能服务

- 多AI引擎：结合DeepSeek、通义千问等AI模型，专门优化HR场景
- 语义理解：准确理解产假政策条款和员工需求的复杂语境
- 会话记忆：记住每次咨询的上下文，提供连贯的政策解答
- 智能推理：基于政策条款自动推算产假天数、津贴标准等

## 🚀 快速开始

### 1. 安装

```
# 克隆仓库
git clone <repository-url>
```

```
cd simple-agent

# 创建虚拟环境
python -m venv venv
source venv/bin/activate # Windows系统: venv\Scripts\activate

# 安装依赖
pip install -r requirements.txt

# 安装额外必需的包
pip install chromadb pypdf unstructured msoffcrypto-tool sentence-transformers
```

## 2. 配置

创建或更新 `config.ini` 文件，添加您的API密钥：

```
[embedding]
api_key = your_qwen_api_key_here
base_url = https://dashscope.aliyuncs.com/compatible-mode/v1/embeddings
model = text-embedding-v4

[deepseek]
api_key = your_deepseek_api_key_here
base_url = https://dashscope.aliyuncs.com/compatible-mode/v1
model = deepseek-v3
```

## 3. 开始聊天！

```
# 启动交互式聊天
python external/call_chatbot.py
```

就这样！您现在可以上传文档并开始提问了。



## 使用示例

# 产假政策咨询

```
> python external/call_chatbot.py
```

HR产假助手：您好！我是专业的HR产假智能助手。  
我可以帮您分析产假政策和员工数据。

您：我想查询北京市的产假政策  
HR产假助手：请上传北京市产假政策文件，或者直接告诉我您想了解的具体问题。

您：data/各地产假规定一览表.xlsx  
HR产假助手：✅ 产假政策文件已加载！  
我可以看到全国各地的产假规定数据。您想了解什么？

您：北京市产假多少天？有什么特殊规定？  
HR产假助手：📄 根据政策文件，北京市产假规定如下：

- 基础产假：98天
- 延长产假：30天
- 总计：128天
- 特殊情况：难产可增加15天...

# 员工产假数据分析

您：我有员工产假申请数据在data/maternity\_leave\_applications.xlsx  
HR产假助手：✅ 员工数据已加载！我可以看到产假申请记录。

您：本年度有多少员工申请了产假？  
HR产假助手：📊 2024年产假申请统计：

- 总申请人数：45人
- 已批准：42人
- 待审核：3人
- 平均产假天数：132天
- 预计总成本：¥890,000

# 批量文档处理

```
# 一次上传多个文档

from external.call_storage import call_storage


# 处理单个文档
result = call_storage("data/report.pdf")
print(result)


# 使用自定义集合名称处理
result = call_storage("data/analysis.xlsx", collection_name="quarterly_reports")


# 不分块处理（适用于较小文件）
result = call_storage("data/summary.txt", chunk_documents=False)


# 使用不同的嵌入提供商
result = call_storage("data/document.pdf", embedding_provider="qwen")
```

## 🔧 可用工具

Simple Agent配备了强大的内置工具：

工具	功能	HR产假应用示例
政策文档检索器	搜索产假政策文档	“查找关于二胎产假的规定”
数据分析器	分析员工产假申请数据	“各部门的产假申请率是多少？”
政策解读器	理解政策条款和上下文	“解释这条产假政策的具体含义”
批量处理工具	处理多地区政策文档	“同时分析全国各省市产假政策”
文件管理器	管理政策和申请文档	“显示所有产假相关文档”

## 支持的文件格式

- ✅ 文档：PDF、DOCX、TXT、Markdown、HTML
- ✅ 数据文件：Excel (.xlsx, .xls)、CSV
- ✅ 批量处理：上传整个文件夹
- ✅ 大文件：每个文件最大100MB

## 配置

### API密钥设置

您需要从这些提供商获取API密钥：

1. 通义千问（阿里云） - 用于嵌入（文本理解）

- 在此获取密钥：[DashScope控制台](#)
- 用途：文档搜索和语义理解

2. **DeepSeek** - 用于AI推理和分析

- 在此获取密钥：[DeepSeek平台](#)
- 用途：聊天回复和数据分析

### 配置文件

使用您的密钥更新 `config.ini`：

```
[embedding]
api_key = your_qwen_api_key
base_url = https://dashscope.aliyuncs.com/compatible-mode/v1/embeddings
model = text-embedding-v4

[deepseek]
api_key = your_deepseek_api_key
base_url = https://dashscope.aliyuncs.com/compatible-mode/v1
model = deepseek-v3
temperature = 0
max_tokens = 2048
```

### 替代方案：环境变量

您也可以将这些设置为环境变量：

```
export QWEN_API_KEY="your_qwen_key"
```

```
export DEEPSEEK_API_KEY="your_deepseek_key"
```

## HR产假管理实际应用场景

### 产假政策咨询与解答

- 政策查询：“上海市二胎产假有多少天？”
- 条款解释：“什么情况下可以申请延长产假？”
- 地区对比：“比较北京、上海、深圳的产假政策差异”
- 特殊情况：“双胞胎、早产、剖腹产的产假如何计算？”

### 员工产假申请管理

- 申请审核：“这位员工的产假申请是否符合政策规定？”
- 天数计算：“根据员工情况，应该批准多少天产假？”
- 费用预算：“本季度产假成本预算是多少？”
- 合规检查：“所有产假申请是否都符合劳动法规定？”

### 产假数据统计与分析

- 趋势分析：“近三年公司产假申请趋势如何？”
- 部门统计：“哪个部门的产假申请最多？”
- 成本分析：“产假津贴支出占人力成本的比例”
- 预测规划：“预计明年需要多少产假预算？”

### 政策文档管理

- 政策更新：快速识别各地产假政策的最新变化
- 文档整理：自动提取和整理各地区产假规定要点
- 合规对照：“公司现行产假制度是否符合最新法规？”

## 故障排除

### 常见问题

"找不到API密钥"错误

- 检查您的 `config.ini` 文件是否包含正确的API密钥
- 验证API密钥是否有效且处于活跃状态（不是虚拟/测试密钥）
- 尝试将环境变量作为替代方案
- 确保通义千问API密钥来自DashScope控制台，而不是占位符

### "找不到文件"错误

- 确保文件路径正确（使用正斜杠）
- 检查文件是否在 `data/` 目录中
- 尝试使用绝对文件路径

### "模型无响应"错误

- 检查您的网络连接
- 验证API密钥是否有足够的额度
- 尝试重启应用程序

### Excel分析不工作

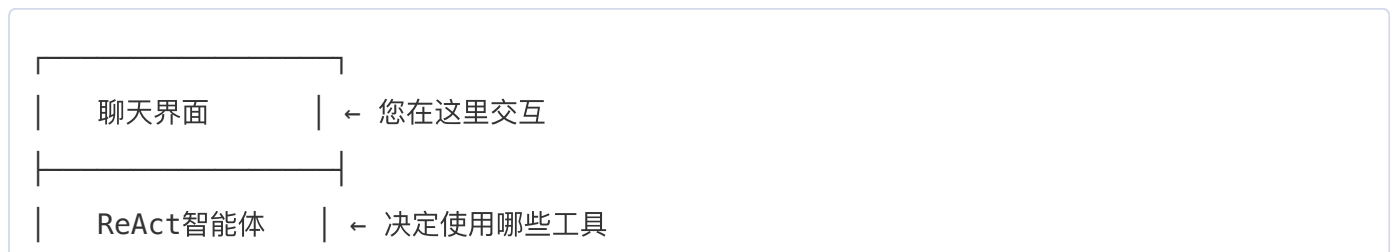
- 确保文件格式为.xlsx、.xls或.csv
- 检查文件大小（最大10,000行，100列）
- 验证文件没有密码保护
- 确保pandas代理具有`allow_dangerous_code=True`（自动处理）
- 检查文件路径解析 - 使用包含空格的确切文件名

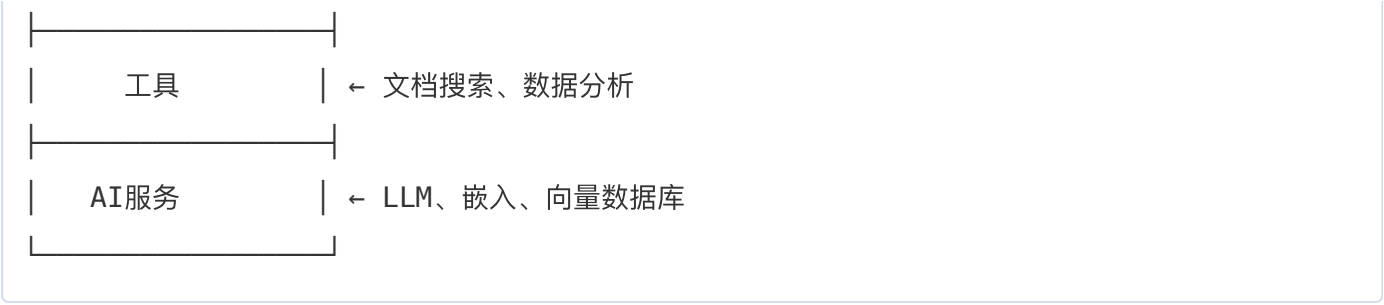
## 获取帮助

- 📖 查看 `tech_documentation.md` 中的技术文档
- 🐛 在仓库中报告问题
- 💬 查看代码注释中的示例用法

## 🏗️ 架构概览

Simple Agent使用模块化架构：





关键组件：

- **ReAct智能体**：选择适当工具的智能推理引擎
- **文档处理器**：处理多种文件格式的智能分块（storage\_utils.py）
- **向量数据库**：ChromaDB与BM25混合搜索，实现快速语义搜索
- **数据分析器**：集成DeepSeek LLM的Pandas驱动Excel/CSV分析
- **记忆系统**：ConversationSummaryBufferMemory维护对话上下文
- **集中配置**：所有服务的通用配置系统