

# 定制化与自动化

推动 R A G 融入企业生产

# 目录

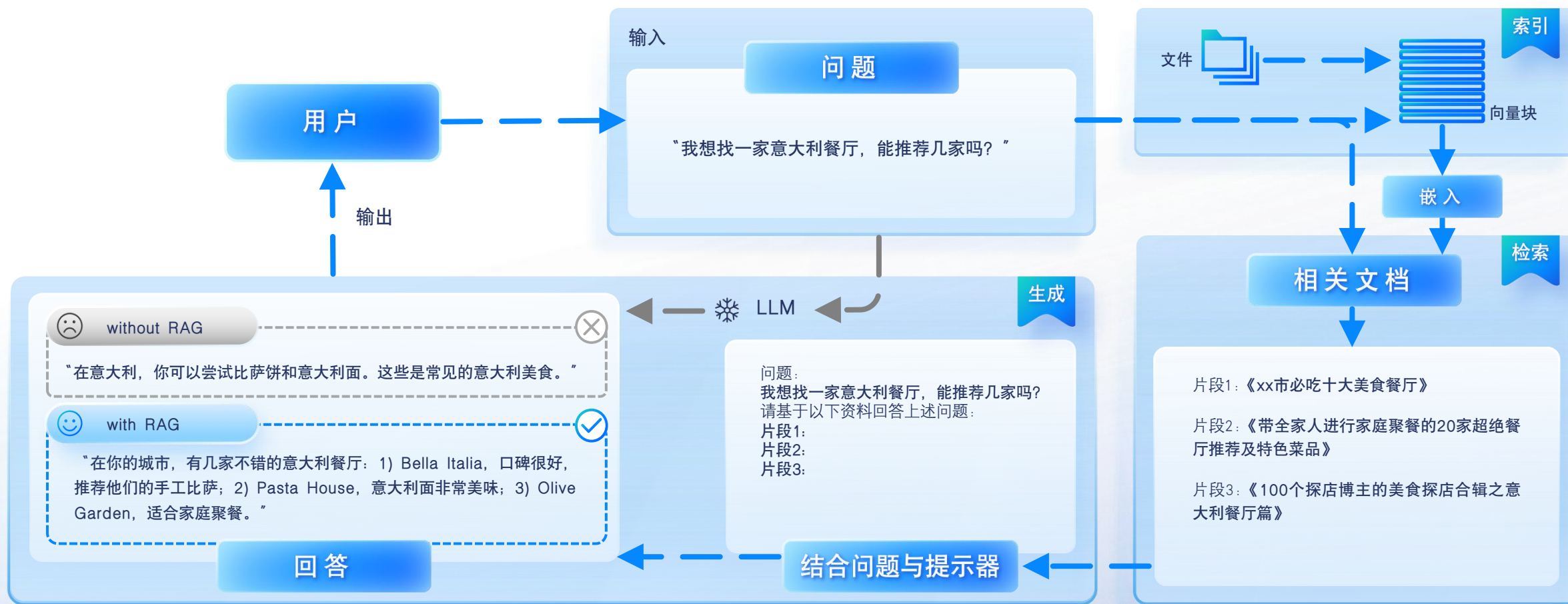
01 传统RAG方案介绍

02 AppBuilder企业级RAG特性详解

03 RAG定制化流程实操

04 RAG自动化流程实操

# 传统RAG方案介绍：技术原理



# 传统RAG方案介绍：解决的问题

大模型信息不足

知识更新缓慢问题

模型幻觉

答案透明度不足问题

...

结合外部最新的知识库，来  
纠正和补充大模型的输出

通过检索和展示与答案相关的  
证据或来源

时效性信息

领域专业信息

私有数据

低成本

快速知识更新方式

提高答案可信度



# AppBuilder企业级RAG特性详解

## 全面企业级功能

### RAG落地难?

RAG知识空间不足

无法适配多种场景

RAG知识储备不足

文字模态场景单一

### 百度智能云千帆 AppBuilder

接入百度云BES和BOS资源，可容纳千亿文件

离线入库、在线检索全流程可调配

融合百度搜索，支持实时全网内容

支持图文解析，检索溯源实现多模态结果生成

无限容量

灵活调配

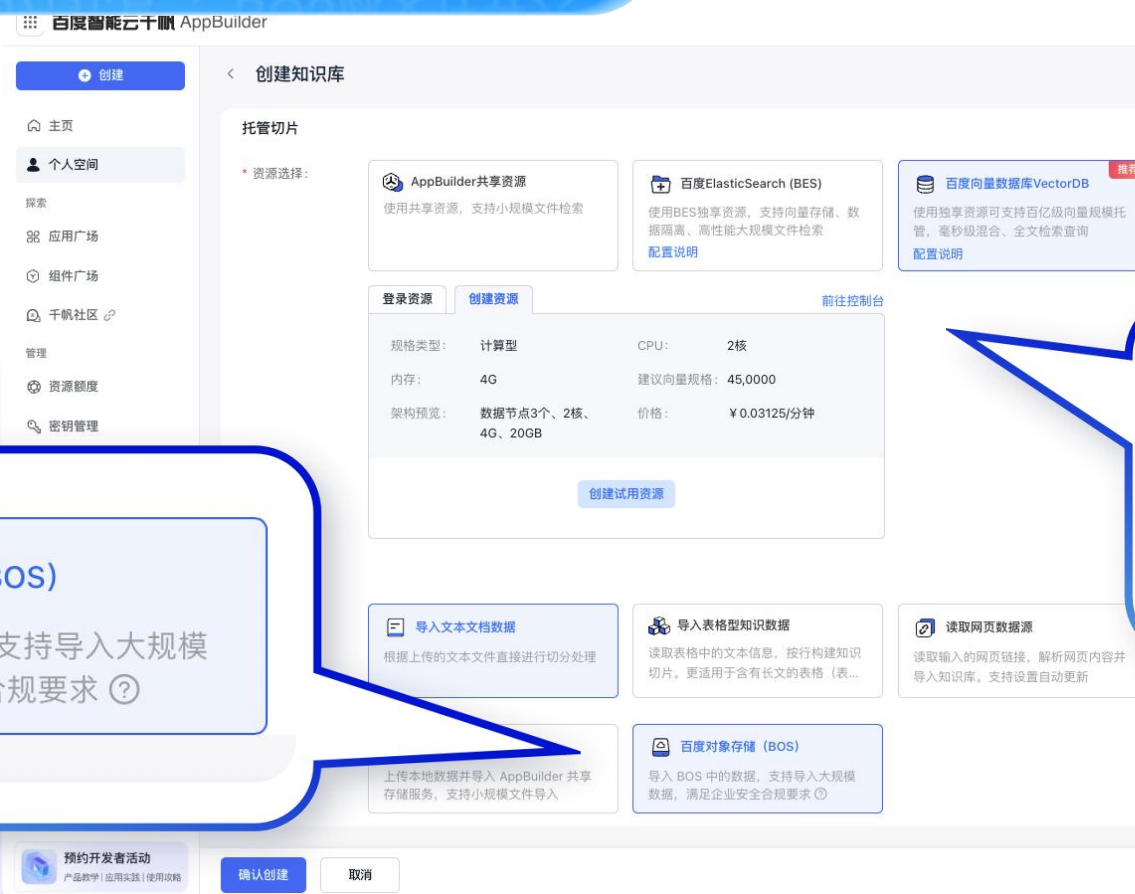
百度AI搜索

多模态问答

# 无限容量：充分利用企业海量业务数据

支持无限容量的知识库存储，实现企业海量文件的RAG，且可根据业务需求灵活扩缩容

打通百度云产品，  
支持BES、VDB切片托管、BOS源文件导入



## 百度对象存储 (BOS)

导入 BOS 中的数据，支持导入大规模数据，满足企业安全合规要求 ②

## 百度向量数据库VectorDB

使用独享资源可支持百亿级向量规模托管，毫秒级混合、全文检索查询

[配置说明](#)

## 百度ElasticSearch (BES)

使用BES独享资源，支持向量存储、数据隔离、高性能大规模文件检索

[配置说明](#)

极致性能与效果

远超同类产品

RAG效果差？

首token响应2~3秒

RAG准确率不高

百度智能云千帆  
AppBuilder

首token响应时间  
最快1.5秒

八大场景准确率  
90%+

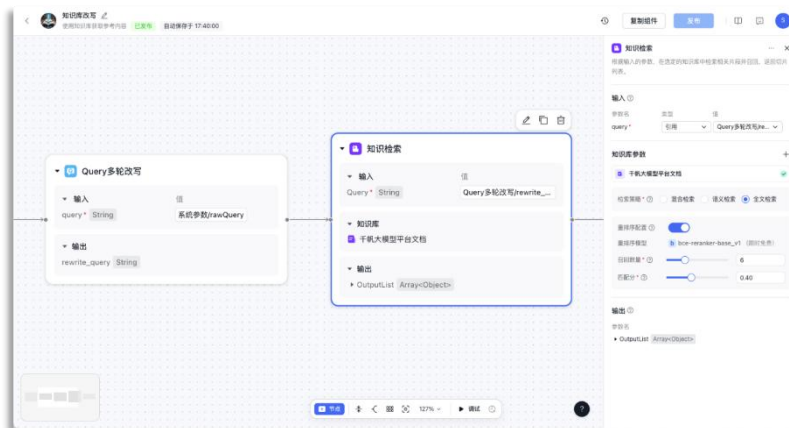
问答效果



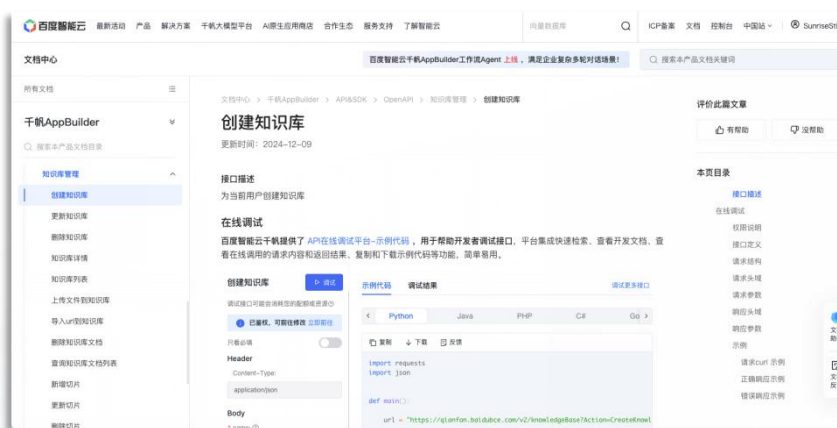


灵活可调

支持灵活配置、调优和被集成



工作流Agent、组件工作流可以原子化编排



知识库检索与增删改查支持API/SDK方式调用



原子能力封装包含解析、切片、增强、检索、各环节能力

百度AI搜索增强

## RAG x 百度搜索

IOT终端



让智能助手拥有  
百度搜索的强大检索能力

高质量数据源及时响应

小型化专家模型准确作答

多模交互的智能体验

安全合规的内容保障

[百度搜索组件：一文读懂多种搜索组合方式](#)

## 多源信息处理

# 解析：图片内容与表格内容

### 配置选择

\* 配置方法：

自定义配置

模版配置

\* 解析策略：

☒ 文字提取

基于规则的文档文字提取

☒ 光学字符识别

图片、扫描件文档需要开启

☒ 版面分析

识别文档文本排版，标题位置信息

☐ 文档图片解析

将图片作为知识库参考信息的一部分

\* 切片策略：

☒ 默认切分

自动设置切分规则

☐ 整文件切片

知识增强：?



☒ 光学字符识别

图片、扫描件文档需要开启

☒ 版面分析

识别文档文本排版，标题位置信息

确认创建

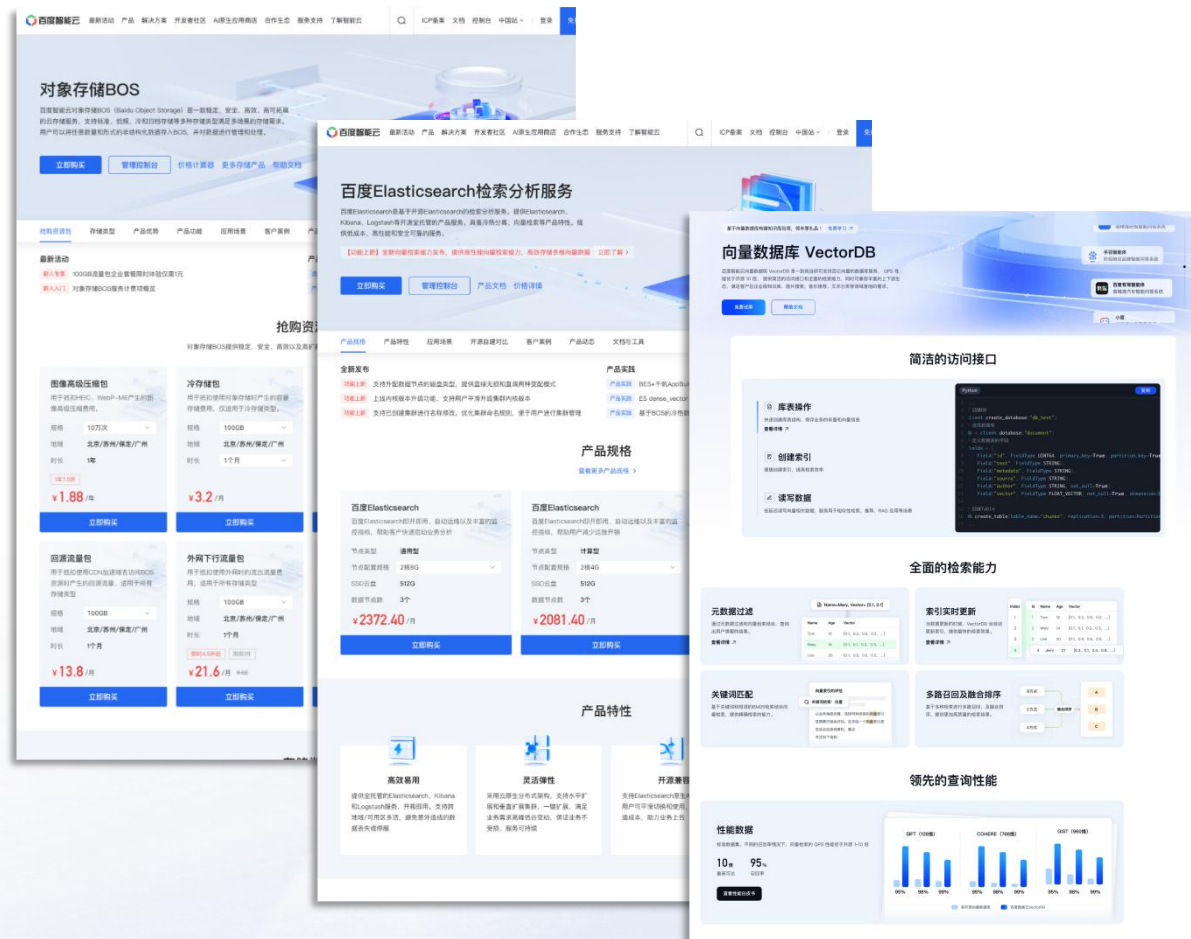
取消

多种解析策略可选择；图片、扫描件、文本排版等



安全可控

满足数据安全合规要求



**可靠:** 支持高可靠的存储和计算, 保障企业级SLA

**安全:** 生产级资源、数据隔离, 且数据内容支持安全过滤

# RAG定制化流程实操

资源存储定制化

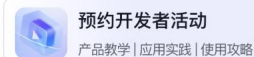
资源导入定制化

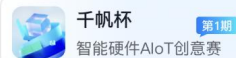
文件类型定制化

文件模板定制化

知识召回定制化

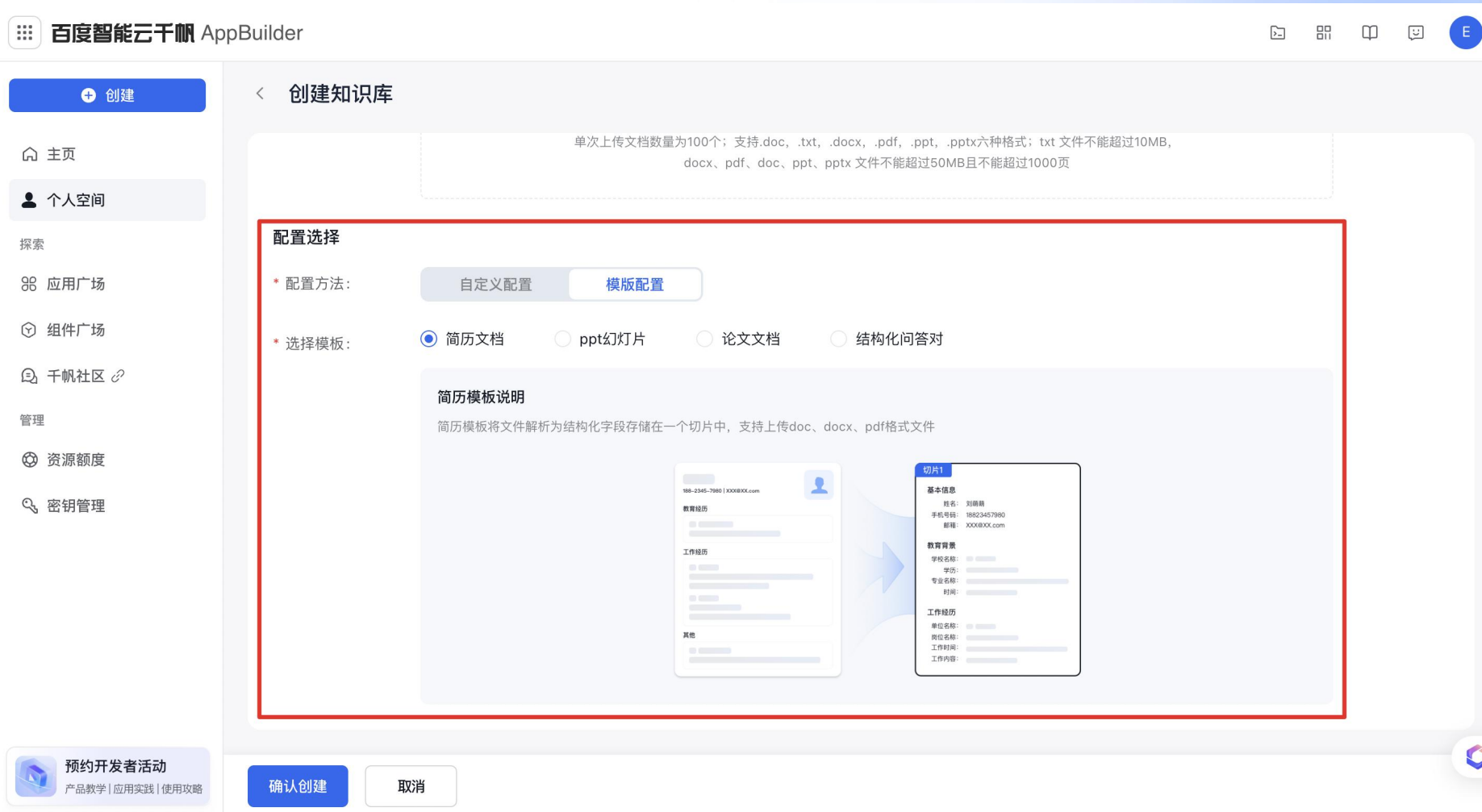
回答干预定制化

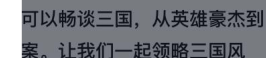












# 回答干预定制化

三国智谈  
带你领略三国风采，解答关于三国...  
自主规划Agent

配置发布调优分析

🏠🗨️C

 数据干预 开

当用户问题匹配已采纳问答对中的问题时，可自动调取已采纳回答输出

 规则干预 敬请期待

当用户问题触发干预规则中的条件时，会自动按照回答规则进行回答

搜索提问/回答关键词 🔍

批量删除 导出 批量处理 添加问答对

待处理 已采纳 不采纳

🚨 用户数据回流需要一定时间，请耐心等待

| <input type="checkbox"/> 提问 | 回答 | 来源 | 编辑时间 | 操作 |
|-----------------------------|----|----|------|----|
|-----------------------------|----|----|------|----|



暂无问答对

通过【添加问答对】可以回流历史对话或上传数据到列表，干预处理后可调优回答效果



# RAG 自动化流程实操

知识库管理自动化

文档管理自动化

知识切片管理管理自动户

RAG应用调用自动化

# 知识库管理自动化

```
import os
import appbuilder
os.environ["APPBUILDER_TOKEN"] = "your_appbuilder_token"

knowledge = appbuilder.KnowledgeBase()
resp = knowledge.create_knowledge_base(
    name="my_knowledge",
    description="my_knowledge",
    type="public",
)
print("新建的知识库ID: ", resp.id)
print("新建的知识库名称: ", resp.name)

# 新建的知识库ID:  da51a988-cbe7-4b24-aa5b-768985e8xxxx
# 新建的知识库名称:  my_knowledge
```

# 文档管理自动化

```
import os
import appbuilder
os.environ["APPBUILDER_TOKEN"] = "your_appbuilder_token"

my_knowledge_base_id = "your_knowledge_base_id"
my_knowledge = appbuilder.KnowledgeBase(my_knowledge_base_id)
print("知识库ID: ", my_knowledge.knowledge_id)

upload_res = my_knowledge.upload_file("./test.txt")
print("文件上传结果: ", upload_res)

add_res = my_knowledge.add_document(content_type='raw_text', file_ids=[
upload_res.id])
print("添加文档结果: ", add_res)

# 知识库ID: da51a988-cbe7-4b24-aa5b-768985e8xxxx
# 文件上传结果: request_id='255eec22-ec87-4564-bdeb-3e5623eaxxxx' id='ef12
119b-d5be-492a-997c-77f8
# 添加文档结果: request_id='412e1630-b570-47c9-a042-caf3cd9dxxxx' knowledg
e_base_id='da51a988-cbe7
```

# 知识切片管理自动化

```
import os
import appbuilder
os.environ["APPBUILDER_TOKEN"] = "your_appbuilder_token"

my_knowledge_base_id = "your_knowledge_base_id"
my_knowledge = appbuilder.KnowledgeBase(my_knowledge_base_id)
print("知识库ID: ", my_knowledge.knowledge_id)
resp = my_knowledge.create_chunk("your_document_id", "content")
print("切片ID: ", resp.id)
chunk_id = resp.id
```





# 课后作业

基础作业

进阶作业

**\*将上述作业在「百度智能云千帆社区」进行发布，发布时选择“千帆大模型训练营”话题**



扫码进入  
百度智能云千帆社区



扫码进入课程群