**知识库训练与模型训练**

**知识库训练**

1. **文件解析与切割**

后台能力支持灵活的文本解析与切割方式：

* **解析方式**：可指定解析类型，比如 PDF 解析、Markdown 解析、自定义正则解析等，还可选择文本编码（如 UTF-8、GBK 等）。
* **切割方式**：包括按字符数、按段落、按句子、自定义符号切割等。
* **参数配置**：可以设置每段的最大字符数、相邻段的重叠字符数，确保语义连续性，适用于嵌入模型的向量化处理。

2. **检索配置**

在知识库训练时，检索配置用于控制如何从知识库中高效、准确地找到相关内容。配置内容主要包括以下几个方面：

* **语法设置（dialect）**：指定所使用的语言或处理规范。
* **相似度阈值（threshold）**：用于筛选召回结果的相关性程度，设定是否启用及阈值大小。
* **重排序设置（rerank）**：用于对初步召回结果进行精细排序，以提升答案的准确性。
* **返回条数（top\_k）**：设置每次查询最多返回的结果数量。
* **向量检索配置（如 Milvus）**：定义索引结构、相似度计算方式、检索参数等底层向量数据库的使用策略，保障查询效率和效果。

这些配置的组合共同决定了知识库检索的性能与准确性，用户可根据应用需求灵活调整。

3. **召回测试**

后台训练时支持手动输入查询内容，对知识库进行问答测试。系统会展示匹配到的片段、匹配分数、对应文件及其位置，便于用户评估知识库的效果。测试支持多轮交互，可帮助优化解析和切割策略。

4. **知识库能力接入数智员工**

训练好的知识库将被作为基础能力层提供给数智员工，在数智员工运行及答案生成时，能快速从知识库中返回高精度匹配的片段

**模型训练**

1. **模型业务选型**

根据业务流程及业务使用场景，分析智能体所需能力，选择最适配的基座模型

2. **模型微调**

根据数智员工工作效果（如政务党务语义理解能力等）对模型权重进行调整，确保数智员工运行能力及效果