

ChromaDB 是一个开源的向量数据库，专门为 AI 应用程序设计。它具有以下特点：

1. 高性能向量检索和相似性搜索
2. 支持多种嵌入模型
3. 简单易用的 API 接口
4. 支持本地部署和云端部署
5. 与 LangChain 等 AI 框架深度集成

DenseRetriever 是 SAGE 框架中的密集向量检索组件，主要功能包括：

1. 支持多种向量数据库后端，包括 ChromaDB 和 FAISS
2. 自动化的文档嵌入和索引
3. 高效的相似性搜索算法
4. 可配置的检索参数（top-k、阈值等）
5. 支持批量检索和实时查询

RAG（Retrieval-Augmented Generation）系统的主要优势：

1. 结合了检索和生成的优点
2. 能够访问外部知识库，提供更准确的答案
3. 减少大语言模型的幻觉问题
4. 支持动态知识更新
5. 提高生成内容的可信度和相关性

向量数据库在 AI 应用中的作用：

1. 存储和管理高维向量数据
2. 提供快速的相似性搜索功能
3. 支持语义搜索和推荐系统
4. 为 RAG 系统提供知识检索能力
5. 实现多模态数据的统一检索

优化检索系统性能的方法：

1. 选择合适的嵌入模型
2. 调整向量维度和索引参数
3. 使用分布式部署提高并发能力
4. 实施缓存策略减少重复计算
5. 优化数据预处理和后处理流程
6. 定期更新和维护向量索引