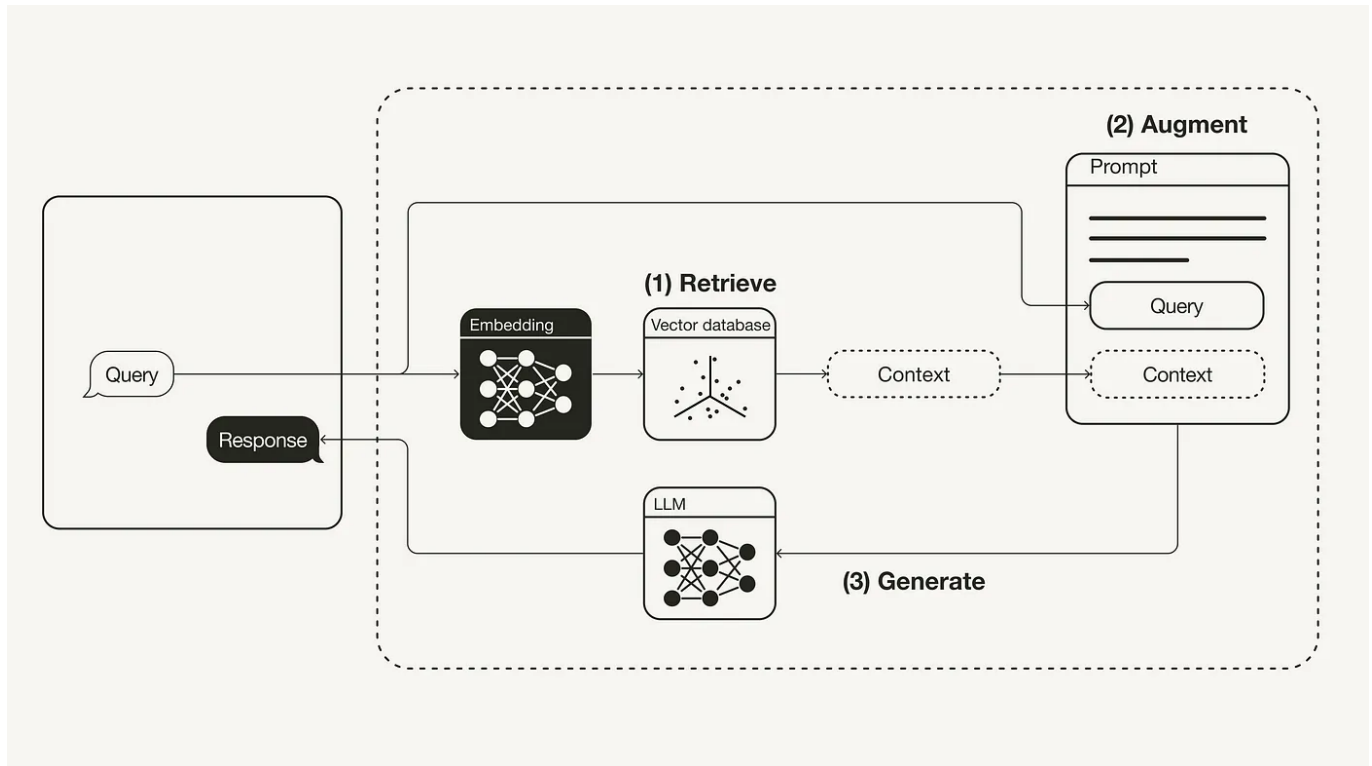


Introduction to RAG System

What is RAG?

Retrieval-Augmented Generation (RAG) is an advanced AI architecture that combines the power of Large Language Models (LLMs) with external knowledge retrieval. This combination allows for more accurate, up-to-date, and contextually relevant responses.



How RAG Works

1. Document Processing Phase



2. Query Processing Phase



3. Response Generation Phase



Project Overview

Our implementation includes:

- Modern web interface
- Real-time processing
- Advanced search capabilities
- Project-based organization
- Comprehensive error handling
- Detailed logging and monitoring

Key Features

The system includes several production-ready features:

- User authentication and authorization
- Multi-format document processing
- Real-time upload progress tracking
- Project-based document organization
- Advanced search capabilities
- Responsive chat interface
- Comprehensive error handling
- Detailed logging and monitoring

Prerequisites

Before starting, ensure you have:

- Basic understanding of PHP and Laravel
- Familiarity with JavaScript
- Access to OpenAI API
- Docker for Elasticsearch
- Basic understanding of vector databases