AI-Powered Email Classification System - Architecture

1. High-Level Architecture

The AI-powered email classification system consists of the following key components:

- **User Interface (Frontend)**: Built using Streamlit to allow users to upload emails and view results.
- **Processing Layer**:
  - **Email Parsing**: Extracts email content and attachments using UnstructuredEmailLoader and PyMuPDFLoader.
  - **Classification & Extraction**: Uses OpenAI's GPT-4 to classify emails and extract key details.
  - **Duplicate Detection**: Employs pgvector to check for semantic similarities.
- **Data Storage (Backend)**:
  - PostgreSQL with pgvector extension for storing processed emails and embeddings.
  - SQLAlchemy with AsyncIO for asynchronous database interactions.

2. System Flow

1. User uploads an email file (EML, PDF, DOCX) via the Streamlit UI.
2. The system extracts email content and attachments.
3. LLM processes the email to classify its request type.
4. The system extracts key details from the email and attachments.
5. Duplicate emails are detected using semantic search.
6. The results are stored in PostgreSQL and displayed in the UI.