Gen AI Orchestrator for Email and Document Triage/Routing

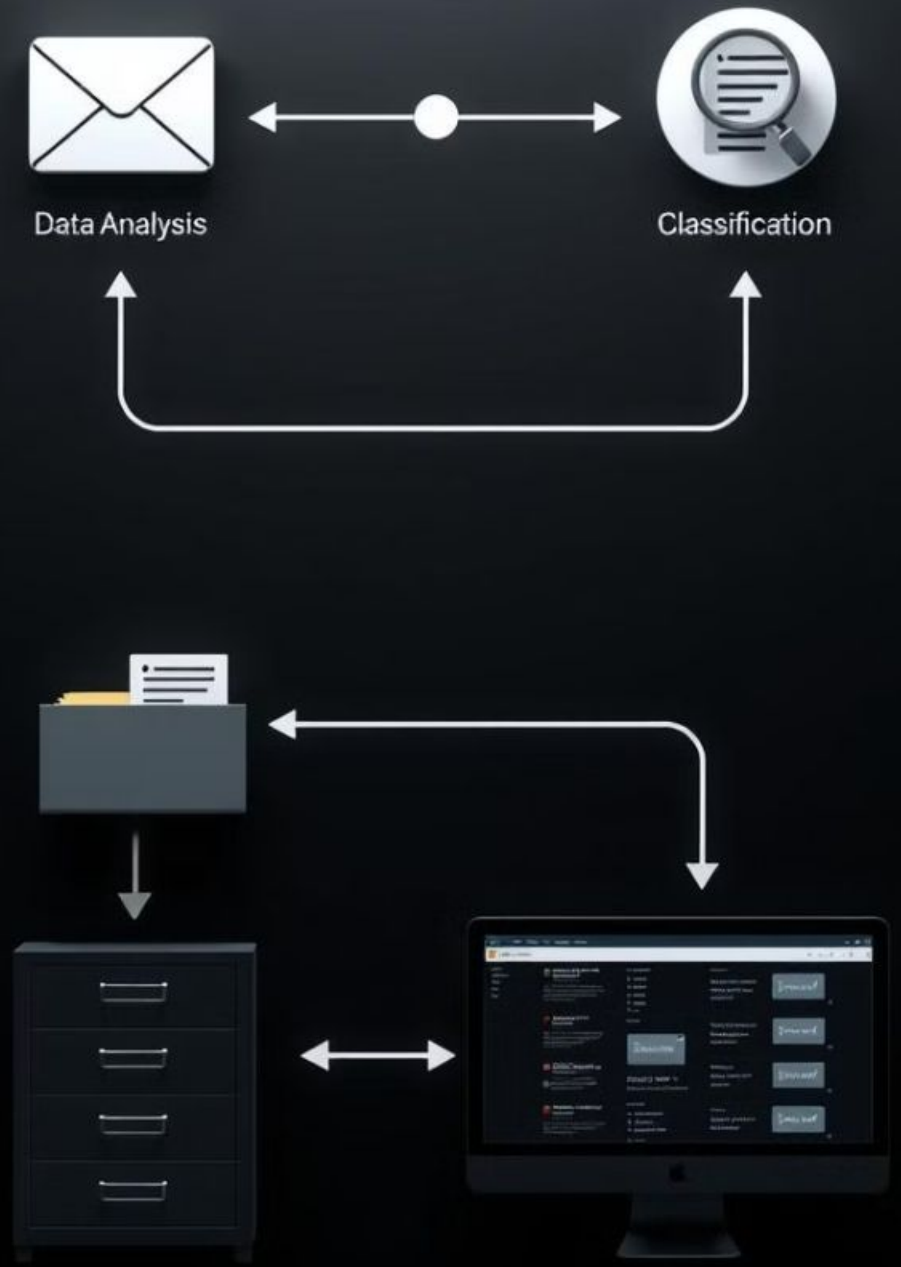
WF I&P Technology Hackathon

Ctrl Alt Defeat

Amith G

Hilmi Parveen

Pebbeti Bhanu Prakash



Core Features

1 Email Classification

Identifies request and sub-request types from email content.

2 Data Extraction

Extracts relevant fields based on the identified request type.

3 Duplicate Detection

Identifies duplicate emails to prevent redundant requests.

4 Multi-Request Handling

Supports emails with multiple request types.

Intelligent Email Processing

Multi-format Support

Processes EML files, PDF email chains, and raw content.

Attachment Processing

Analyzes content from PDF, Word, HTML, text, and images.

HTML Extraction

Cleans and extracts text from HTML email bodies.

Why It Stands Out

- Request Type Identification

  Classifies emails into primary and secondary request types after reading the email content and attachments.

- Sub-request Classification

  Identifies specific sub-categories within each request type.

- Intelligent Duplicate Detection

  Looks for semantic similarity between requests to identify duplicate requests and avoid redundant operations.

- Fully Configurable

  The request-types and sub-request types are fully configurable from the UI.

- Attribute Extraction

  Extracts the relevant data from the e-mail and attachments needed to process the request

- Priority Detection

  Determines the primary intent when emails contain multiple requests.

Intelligent Duplicate Detection

Semantic Similarity

Detects duplicates using embeddings, even with wording variations.

Metadata Comparison

Uses sender, recipient, thread ID, and IP address for detection.

Confidence Scoring

Provides granular confidence levels for duplicate detection.

Data Extraction

1

Field Extraction

Extracts structured data like amounts and account numbers.

2

Source Prioritization

Prioritizes data sources based on field type.

3

Confidence Scoring

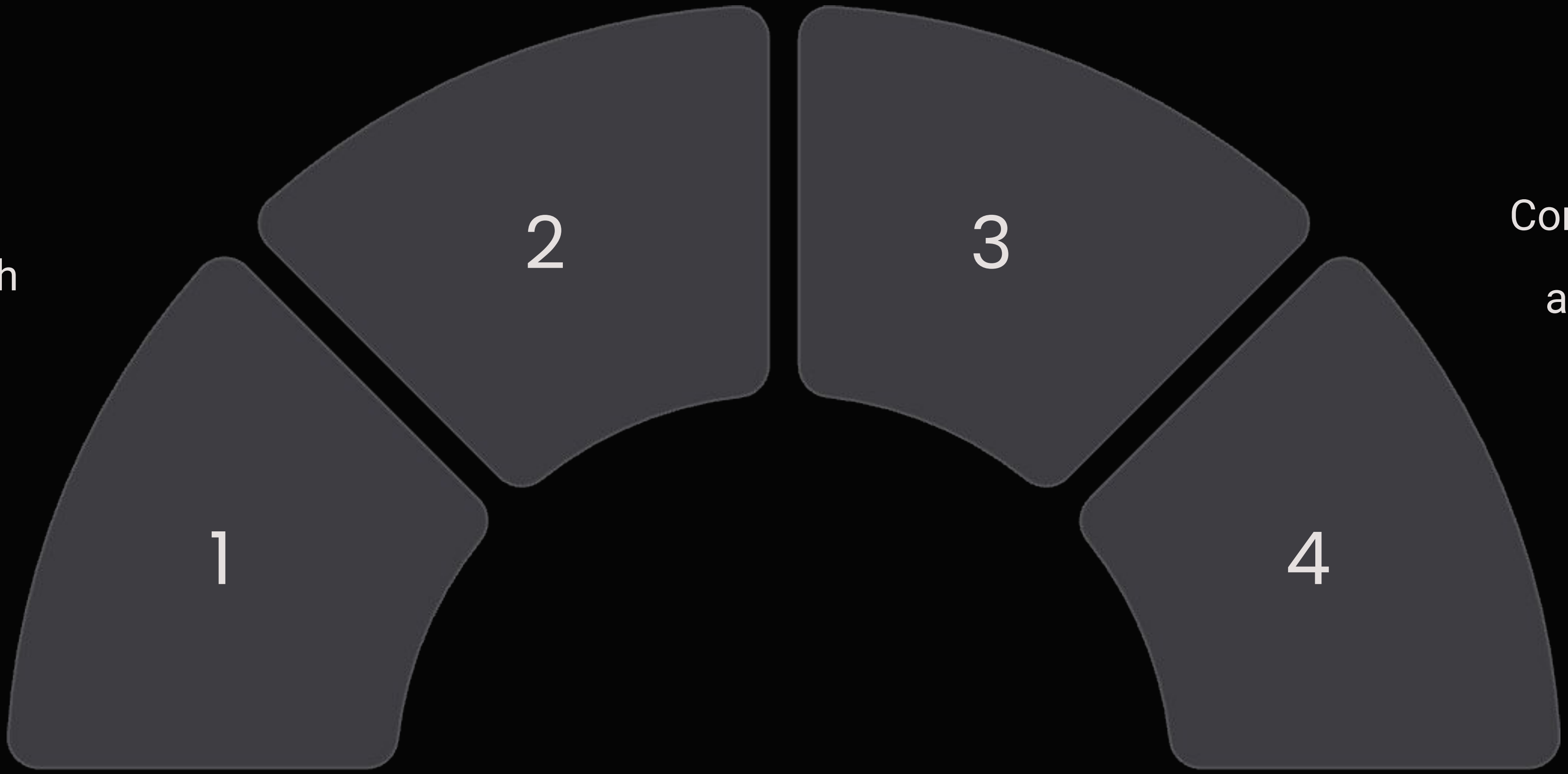
Assigns confidence levels to extracted values.

4

Format Normalization

Standardizes dates and currency values.

Robust Architecture



API Key Management

Rate-limited API keys with automatic rotation.

Multi-LLM Support

Flexible integration with various LLM providers.

Task-specific LLM Routing

Uses different models for classification vs. data extraction.

Error Handling

Comprehensive error handling and fallback mechanisms.

Key Components

1

EmailProcessor

Extracts text and metadata from email content.

2

IntelligentDuplicateDetector

Identifies duplicate emails using semantic similarity.

3

ClassificationService

Orchestrates the email classification workflow.

API Endpoints

POST /classify-email-chain	Process email chain from PDF file.
POST /classify-eml	Process email from an EML file.
GET /request-types	Retrieve all request types and sub-types.

Configuration Options

14

Duplicate cache days (default).

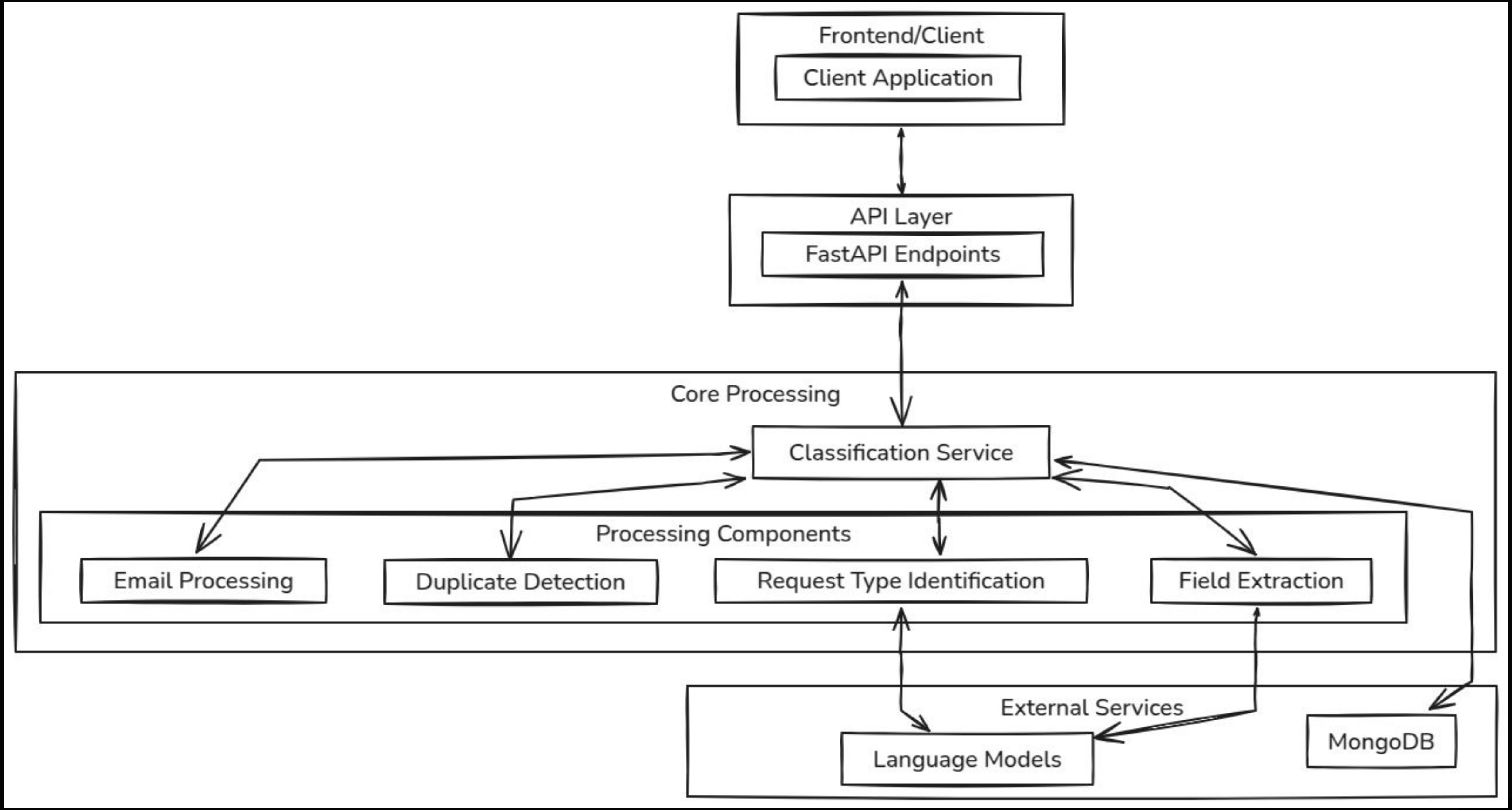
0.8

Semantic threshold (default).

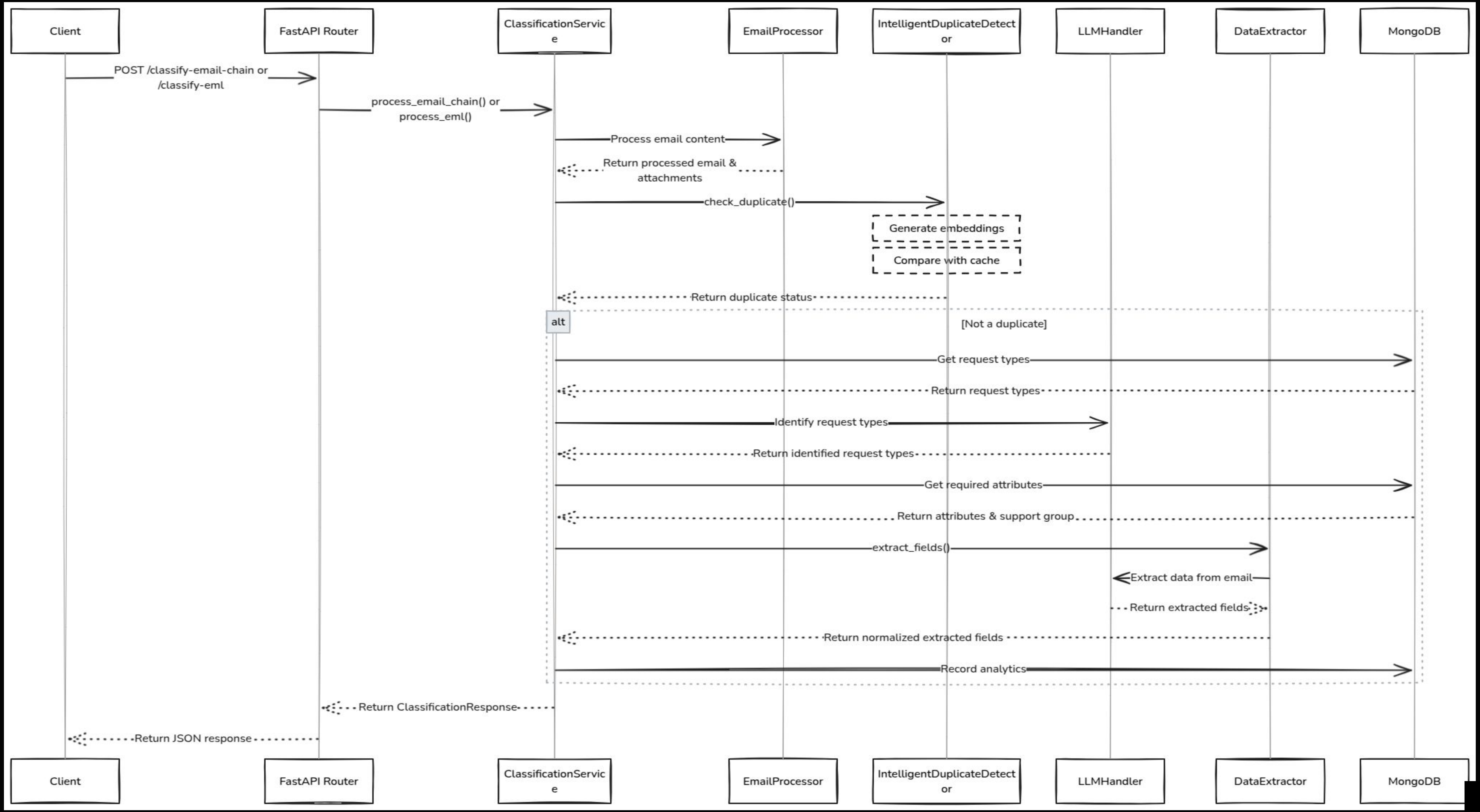
10

Max attachment size MB (default).

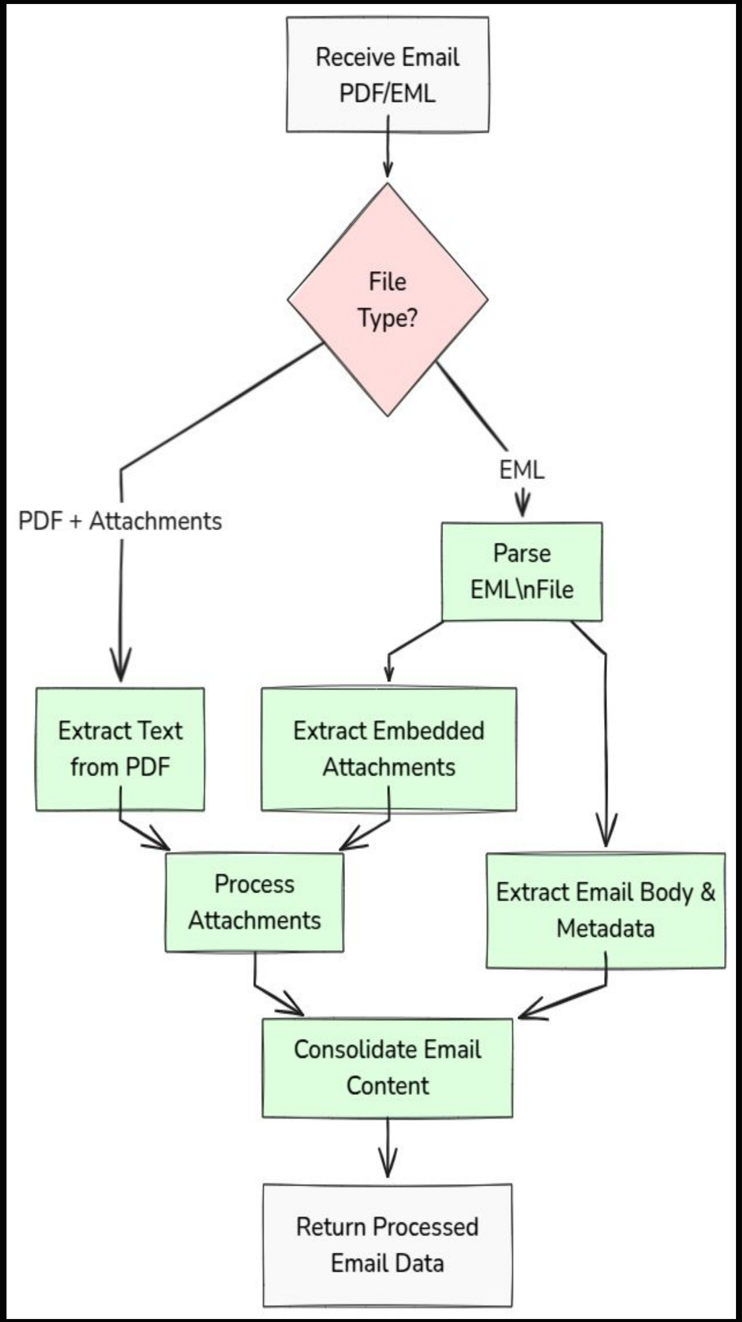
Overall high level view

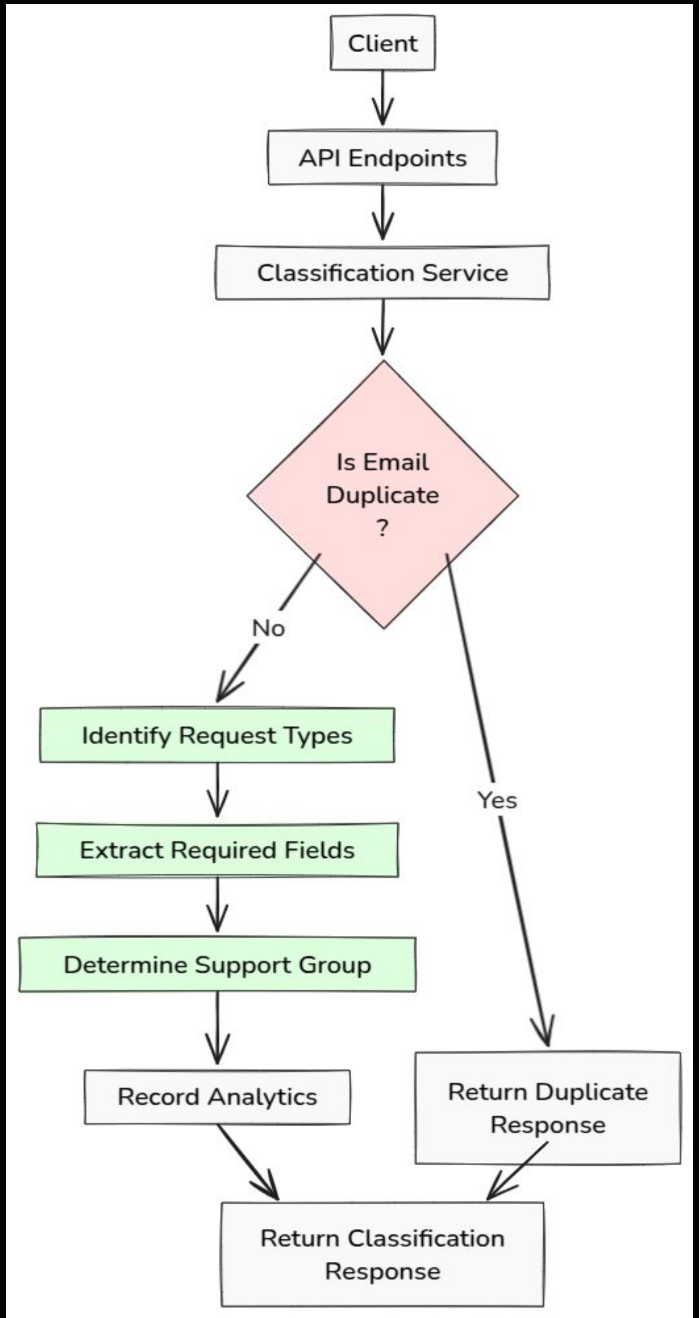


Overall flow diagram

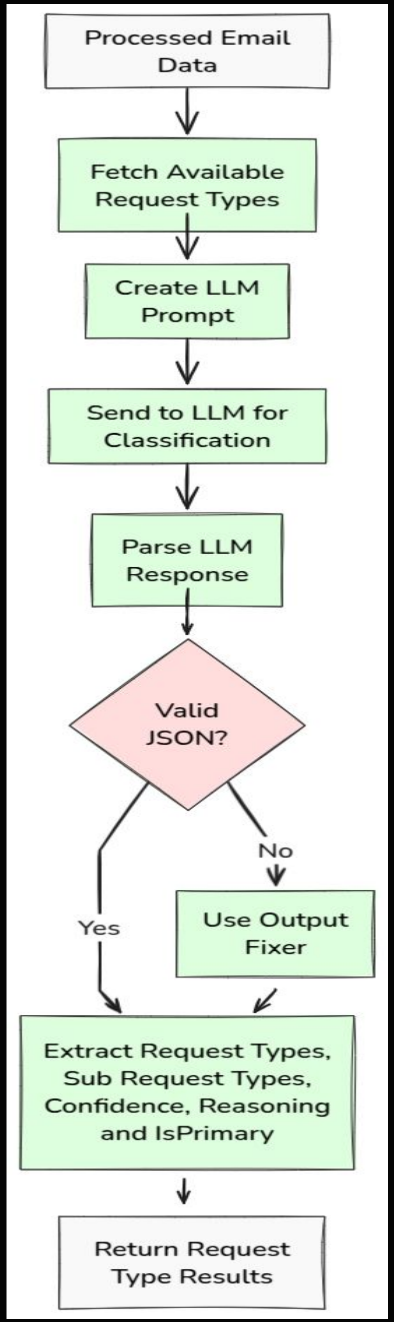


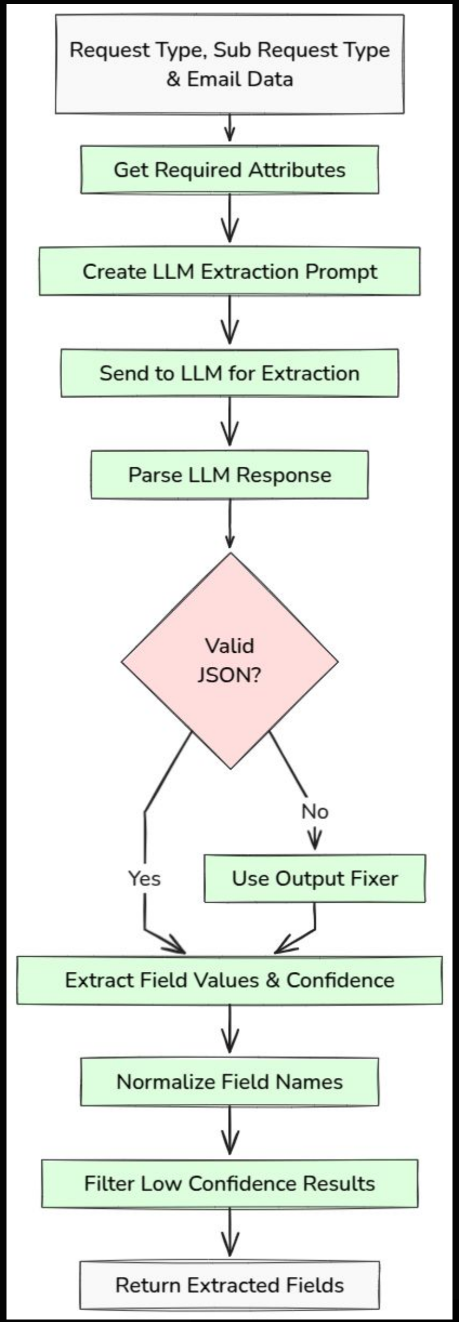
Email processing and intelligent duplicate detection flow


Receive Email PDF/EML
File Type?
PDF + Attachments
EML
Parse EML\nFile
Extract Text from PDF
Extract Embedded Attachments
Process Attachments
Extract Email Body & Metadata
Consolidate Email Content
Return Processed Email Data



Request type and fields extraction flow


Processed Email Data
Fetch Available Request Types
Create LLM Prompt
Send to LLM for Classification
Parse LLM Response
Valid JSON?
Yes
No
Use Output Fixer
Extract Request Types, Sub Request Types, Confidence, Reasoning and IsPrimary
Return Request Type Results


Request Type, Sub Request Type & Email Data
Get Required Attributes
Create LLM Extraction Prompt
Send to LLM for Extraction
Parse LLM Response
Valid JSON?
Yes
No
Use Output Fixer
Extract Field Values & Confidence
Normalize Field Names
Filter Low Confidence Results
Return Extracted Fields

Future Scope

1 Integration with live email servers

Classify emails as it comes in.

2 Integration with a ticketing system

Automate the user story creation based on support groups.

3 Batch processing

Process emails as a batch using multi-thread.

4 Manual review tags.

Give tags for manual review if model confidence is low.