

BRD Analysis - 2025-08-08

Okay, here's a summary of the key requirements extracted from the provided BRD content:

- Core Functionality & API Overview:**
 - AI-Powered Document Processing:** The API enables users to analyze various document types (PDF, Excel, images, videos) using AI.
 - RAG Implementation:** Supports Retrieval-Augmented Generation for question answering based on document content.
 - Document Classification:** Can classify document types (e.g., Commercial Invoice, Bill of Lading).
 - Supported Document Types:** PDF, Excel (XLS, XLSX), Images (JPG, PNG, JPEG), Videos (MP4).
 - AI Model Options & Processing:**
 - Multiple AI Model Choices:** Supports `gemma3:12b`, `gpt-4.1-mini`, `simplyfillm`, `llama3.2-vision:latest`, and `mistral:latest` to tailor response quality.
 - simplyfillm – Default:** Runs all models and selects the best response.
 - Chunking & Relevant Context Retrieval (RAG):** Utilizes document chunking and context retrieval for RAG analysis.
 - Request & Response Structure:**
 - POST Method:** All requests utilize the POST method to the `/file` endpoint.
 - multipart/form-data:** Requests are submitted using `multipart/form-data`.
 - Request Parameters:** Uses the `content` parameter (text query) and file upload parameters.
 - Structured Response:** The API returns a JSON response including:
 - description:** Text content generated by the selected model.
 - audit_trail:** Detailed log of the request, including timestamps, request IDs, metrics, and event logs (input received, guardrail checks, errors).
 - guardrail_info:** Details of guardrail checks (verdict, classification, explanation).
 - Guardrails & Error Handling:**
 - Content Filtering (Guardrails):** Includes profanity detection and scam/fraud detection.
 - 403 Error:** Returned when content is flagged by guardrails.
 - 500 Error:** Indicates an internal server error, likely related to file format issues.
 - Common Error Codes:** 400 (missing prompt), 500 (server error).
 - File Guidelines:**
 - PDFs:** No size limits, all pages processed.
 - Excel:** XLS and XLSX formats supported.
 - Images:** JPG, PNG, JPEG formats; 300 DPI recommended.
 - Videos:** MP4 format recommended, maximum 5 minutes duration.

Do you want me to elaborate on any of these points, focus on a specific aspect (e.g., the guardrail system, a particular use case, or the data format requirements), or would you like me to create a diagram (e.g., a flow chart) illustrating the API's processing flow?