# BRD Analysis - 2025-08-08

Okay, let's analyze the information provided, focusing on extracting key requirements and identifying potential gaps, given the context of the IDFC Gen-AI App Store BRD and the supplementary summary about Supradeep Reddy Mabbu. **Overall Assessment & Initial Observations:** The provided snippets – the 10-point BRD summary and the subsequent information about Supradeep – paint a picture of a sophisticated AI application focused on document processing and automation within the financial sector. The core value proposition is leveraging Gen-AI to extract insights from various document types (Commercial Invoices, Bills of Lading, etc.). Supradeep's work reinforces this, showcasing practical application of AI agents and technical skills. **Key Requirements Extracted:** 1. **Document Classification:** The system *must* support document classification, specifically mentioning "Commercial Invoice" and "Bill of Lading" as examples. This implies a classification engine is a core component. *Missing Information:* The BRD doesn't detail the classification accuracy requirements, the number of document types the system will support, or the training data used for the classification models. 2. **RAG Process:** The RAG process is fundamental. It involves: * **Chunking:** Breaking down documents into manageable segments. * **Relevance Identification:** Determining the most pertinent chunks for answering a user query. * **Answer Generation:** Synthesizing information from the selected chunks to create a response. *Missing Information:* The BRD lacks specifics on the chunking strategy (e.g., fixed size, semantic chunking), the algorithm for relevance identification, and the methods used for generating answers (e.g., template-based, generative models). 3. **Content Parameter Requirement:** The "content" parameter is *required* for all user queries. This suggests a standardized input format is necessary for the API to function correctly. *Missing Information:* The BRD doesn't define the structure or data types expected within the "content" parameter. 4. **AI Model Flexibility:** The ability to select from various AI models (gemma3:12b, gemma3:4b, GPT-4.1-mini, etc.) highlights a key design principle – optimizing for accuracy and speed. *Missing Information:* The BRD needs to define the criteria for selecting an AI model (e.g., based on document type, complexity of the query, desired latency). It also needs to specify the infrastructure requirements for running these models. 5. **Guardrails:** The system incorporates guardrails for preventing sensitive content generation. *Missing Information:* It needs to specify the type of content that's considered "sensitive" and how these guardrails are implemented (e.g., content filtering, model prompting techniques). **Potential Gaps & Questions for Clarification:** * **Document Volume & Velocity:** What is the expected volume of documents the API will process, and at what rate? This will impact infrastructure and performance requirements. * **Accuracy Targets:** What are the acceptable accuracy levels for document classification and RAG? * **Error Handling:** The BRD mentions error rates, but it's crucial to define specific error handling strategies (e.g., retry mechanisms, fallback responses). * **Scalability:** How will the API scale to handle increased demand? * **Security:** What security measures are in place to protect sensitive data? **Next Steps:** To fully analyze this information, we need the complete BRD. However, based on these snippets, I've identified key requirements and potential gaps. I would recommend focusing initial efforts on clarifying the accuracy targets, the RAG chunking strategy, and the security considerations. Do you want me to: * Generate a high-level architecture diagram based on these snippets? * Create a simplified flowchart illustrating the RAG process? * Focus on a specific aspect of the BRD (e.g., the AI model selection process)?