Product Requirements Document

Project: BFSI Call Center AI Assistant

1. Objective and Scope

The objective of this project is to build a lightweight, compliant, and efficient AI assistant for handling common Banking, Financial Services, and Insurance (BFSI) call center queries. The assistant should provide fast, accurate, and standardized responses.

The assistant will support queries related to:

• Loan eligibility and application status

• EMI details and schedules

• Interest rate information

• Payment and transaction queries

• Basic account and customer support

Key System Mandates:

• Must run locally using a small language model

• Must prioritize responses from a curated dataset

• Must use retrieval for complex financial queries

• Must follow strict safety and compliance guidelines

2. Technical Stack and Architecture Flow

System Pipeline: User Query → Alpaca Dataset Similarity Check → Local Fine-Tuned SLM → RAG Layer (if required) → Final Response

The system is designed to rely primarily on dataset-based responses to ensure safety and compliance.

3. Core System Components

3.1 Dataset (Primary Response Layer)

• Minimum 150+ BFSI conversation samples

• Format: Alpaca (Instruction, Input, Output)

• Professional and compliant tone

Role: If a strong similarity match is found, return the stored response directly.

3.2 Small Language Model (Local SLM)

• Lightweight instruction-based model

• Fine-tuned using the Alpaca dataset

• Runs locally on modest hardware

Role: Used only when no strong dataset match is found.

3.3 RAG Layer (Knowledge Retrieval)

• Handles complex financial or policy-related queries

• Used for interest explanations, EMI breakdown, penalties, policy rules

• Knowledge stored in structured documents

Role: Retrieve relevant information and generate grounded responses for complex queries.

4. Response Logic (Priority Order)

Tier 1: Dataset Match – Return stored response if similarity is strong.

Tier 2: Fine-Tuned SLM – Generate response if no dataset match.

Tier 3: RAG Retrieval – Use knowledge documents for complex queries.

5. Guardrails and Security

• No guessing of financial numbers

• No generation of fake rates or policies

• No exposure of sensitive customer data

• Reject out-of-domain or unsafe queries

• Maintain regulatory compliance at all times

6. Scalability and Maintainability

• System should support scaling for higher call volumes

• Dataset, model weights, and RAG documents must be version controlled

• Easy updates to policies and financial documents

7. Deliverables

• 150+ Alpaca formatted BFSI dataset

• Fine-tuned small local model (SLM weights)

- Structured knowledge base for RAG
- Working end-to-end demo
- Technical documentation explaining architecture and logic