**IDFC Gen-AI App Store Documentation**

**Overview**

The IDFC Gen-AI App Store is a comprehensive AI-powered solution for banking and financial document processing. This API enables users to interact with various document types including PDFs, Excel sheets, images, videos, and perform advanced operations like Retrieval-Augmented Generation (RAG) analysis and document classification.

**API Endpoints**

**Health Check**   
Endpoint:GET /health   
Description: Returns the current health status of the service along with performance metrics.

Response:

{   
 "status": "healthy",  
 "service": "IDFC Gen-AI App Store API",   
 "metrics": {  
 "total\_requests": 150,  
 "total\_errors": 5,  
 "guardrail\_blocks": 2,  
 "success\_rate": "96.7%"  
 }

}

**Metrics**

Endpoint:GET /metrics   
Description: Provides detailed monitoring metrics for the API service.

Response:   
{   
"metrics": {   
"total\_requests": 150,   
 "total\_errors": 5,   
"guardrail\_blocks": 2,   
"success\_rate": "96.7%",   
 "error\_rate": "3.3%",   
"guardrail\_block\_rate": "1.3%"   
 }   
}

**Chat Interface**   
Endpoint:POST/ http://localhost:9080/chat   
Method: POST   
Content-Type: multipart/form-data   
**Request Parameters**   
**Form Data Parameters**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PARAMETER | TYPE | REQUIRED | DESCRIPTION | EXAMPLE |
| content | string | Yes | The user query or message | "What is the total amount in the invoice?" |
| model | string | No | AI model to use (default: gemma3:12b) | "gemma3:12b", "gpt-4.1-mini", "simplyfillm" |

**File Upload Parameters**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| FIELD NAME | TYPE | REQUIRED | DESCRIPTION | SUPPORTED FORMATS |
| pdf | file | No | Standard document analysis | PDF |
| excel | file | No | Spreadsheet analysis | XLS, XLSX |
| image | file | No | Image document analysis | JPG, PNG, JPEG |
| video | file | No | Video frame analysis | MP4 |
| rag | file | No | RAG-based document Q&A | PDF |

**Supported AI Models**

|  |  |  |
| --- | --- | --- |
| MODEL | DESCRIPTION | BEST FOR |
| simplyfillm | Runs all models in parallel and selects the best response | Highest quality responses, complex queries |
| gpt-4.1-mini | Advanced reasoning and complex tasks | Detailed analysis, complex financial queries |
| gemma3:4b | Quick queries and responses | Simple questions, fast responses |
| gemma3:12b | Enhanced reasoning capabilities | General purpose, default option |
| llama3.2- vision:latest | Advanced visual document analysis | Image and document analysis |

|  |  |  |
| --- | --- | --- |
| mistral:latest | Specialized financial expertise | Financial data analysis, banking procedures |

**Usage Scenarios**   
**1. Document Analysis (PDF)**   
Scenario: Analyzing a standard PDF document like a bank statement or financial report.

Request:

|  |  |
| --- | --- |
| • | Form Data: content="Analyze this bank statement" |

•File: Upload PDF using the pdf field   
•Model: gemma3:12b (default)   
Response: Text content extracted from the PDF with analysis of key information.

**2. Retrieval-Augmented Generation (RAG)**   
Scenario: Asking specific questions about a document's content.

Request:

|  |  |
| --- | --- |
| • | Form Data: content="What is the total amount due?" and model="gpt-4.1-mini" |

•File: Upload PDF using the rag field   
Process:   
1.Document is chunked into smaller segments   
2.Relevant chunks are identified based on the query   
3.Answer is generated using context from relevant chunks Best For: Extracting specific information from lengthy documents.

**3. Document Classification**   
Scenario: Identifying types of documents in a multi-page PDF.

Request:

|  |  |
| --- | --- |
| • | Form Data: content="What documents are present in this file?" |

•File: Upload PDF using the pdf field   
Supported Document Types:

•Commercial Invoice   
•Bill of Lading   
•CRL document   
•Bill of Entry   
•Others

Response: Classification of each page and summary of document types found.

**4. Excel Analysis**

Scenario: Analyzing financial data in spreadsheets.

Request:

|  |  |
| --- | --- |
| • | Form Data: content="Analyze the financial trends in this spreadsheet" |

•File: Upload Excel file using the excel field

Response: Formatted text representation of the spreadsheet data with analysis.

**5. Image Analysis**

Scenario: Extracting information from images of checks, forms, or documents.

Request:

|  |  |
| --- | --- |
| • | Form Data: content="What is the amount on this invoice?" |

•File: Upload image using the image field

Response: Extracted text and specific information requested from the image.

**6. Video Analysis**

Scenario: Analyzing banking procedures or verification videos.

Request:

|  |  |
| --- | --- |
| • | Form Data: content="Describe the procedure shown in this video" |

•File: Upload video using the video field (MP4 format recommended)

Process:

•Key frames are extracted at 1-second intervals •Up to 5 key frames are analyzed   
•Video metadata is provided

Response: Description of the video content based on analyzed frames.

**Response Format**

{   
 "description": "Response from gemma3:12b:\nThe total amount due is ₹50,000.", "audit\_trail": {   
 "timestamp": "2024-01-15 10:30:45",   
 "request\_id": "a1b2c3d4-e5f6-7890-g1h2-i3j4k5l6m7n8",   
 "metrics": {   
 "total\_requests": 150,   
 "error\_count": 5,   
 "guardrail\_blocks": 2,   
 "response\_time": 2.345,   
 "success\_rate": 96.7   
 },   
 "events": [   
 {   
 "timestamp": "2024-01-15 10:30:45",   
 "event": "input\_received",   
 "details": {   
 "model": "gemma3:12b",   
 "content\_length": 45   
 }   
 },   
 {   
 "timestamp": "2024-01-15 10:30:45",   
 "event": "guardrail\_check",

"details": {   
 "verdict": "no",   
 "classification": "safe",   
 "explanation": "No sensitive content detected" }   
 }   
 ]   
 },   
 "guardrail\_info": {   
 "verdict": "no",   
 "classification": "safe",   
 "explanation": "No sensitive content detected" }   
}

**Error Handling**

**Common Errors**

|  |  |  |  |
| --- | --- | --- | --- |
| STATUS CODE | ERROR | DESCRIPTION | SOLUTION |
| 400 | No prompt provided | Request missing text query | Add a 'content' parameter with your question |
| 403 | Input  flagged by guardrail | Content detected as potentially harmful | Revise query to remove flagged content |

|  |  |  |  |
| --- | --- | --- | --- |
| 500 | Error  processing request | Internal server error | Check file format and try again |

**Error Response**

{   
 "error": "No input provided.",   
 "audit\_trail": {   
 "timestamp": "2024-01-15 10:30:45",   
 "request\_id": "a1b2c3d4-e5f6-7890-g1h2-i3j4k5l6m7n8", "metrics": {   
 "total\_requests": 150,   
 "error\_count": 5,   
 "guardrail\_blocks": 2   
 },   
 "events": [   
 {   
 "timestamp": "2024-01-15 10:30:45",   
 "event": "error",   
 "details": "No input provided"   
 }   
 ]   
 }   
}

**Security Features**

**Guardrail System**

The API includes a content filtering system that analyzes input for potentially harmful content:

•Profanity Detection: Flags inappropriate language   
•Scam/Fraud Detection: Identifies potential fraudulent content •Custom Rules: Additional rules can be configured

When content is flagged, the request is blocked with a 403 response.

**Best Practices**

**For Optimal Results**

1.Use simplyfillm for critical queries - It runs all models and selects the best response 2.Be specific in your questions - Clear, focused queries yield better results   
3.Use RAG for document Q&A - Upload documents with the 'rag' key for detailed analysis   
4.Classify documents first - Use document identification queries to understand your files

**File Guidelines**

•PDFs: No size limit, 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

**Getting Started**

**Sample cURL Commands**

Basic Text Query:

curl -X POST http://localhost:9080/chat \

-F "content=What banking services does IDFC offer?"

PDF Analysis:

curl -X POST http://localhost:9080/chat \  
 -F "content=Analyze this bank statement" \   
 -F "pdf=@statement.pdf" \   
 -F "model=gemma3:12b"

RAG Question Answering:

curl -X POST http://localhost:9080/chat \  
 -F "content=What is the invoice total amount?" \ -F "rag=@invoice.pdf" \   
 -F "model=gpt-4.1-mini"

Document Classification:

curl -X POST http://localhost:9080/chat \   
 -F "content=What documents are present in this file?" \ -F "pdf=@documents.pdf"