Business Problem, Target Audience, and Core Features

Business Problem

Issue

Many businesses rely on automated document processing tools for invoices, purchase orders, and financial records. Traditional document parsing techniques, such as regex-based extraction, often struggle with:

- **Unstructured or inconsistent document formats** – Different suppliers or clients use varied invoice layouts.
- **Human errors in data entry** – Misspellings, missing fields, or incorrect placements can break regex-based extraction.
- **Lack of real-time resolution** – Users must manually correct missing or incorrect fields, leading to inefficiencies.

Solution

To overcome these issues, the **AI-based document support agent** will:

- **Extract data efficiently** using regex and AI-based OCR methods.
- **Identify missing fields** and proactively flag incomplete data.
- **Use AI-driven predictions** to infer missing data based on document context.
- **Provide an interactive AI chatbot** that allows users to resolve data gaps through a conversational interface.

This **streamlined, intelligent document processing workflow** will significantly reduce manual corrections, improving accuracy and efficiency.

Target Audience

The solution is designed for industries and teams that rely on document-based workflows, such as:

1. **Logistics & Supply Chain Management**
   - Companies handling invoices, shipment documents, and customs paperwork.
2. **Finance & Accounting Departments**
   - Teams processing expense reports, invoices, and tax documents.
3. **Legal & Compliance Teams**
   - Businesses that require structured contract analysis and compliance checks.
4. **SaaS-Based Document Processing Tools**
   - Organizations looking to enhance existing document-processing platforms like **Go Invoice** with AI-driven automation.

Core Features

1. Document Upload & Processing

✅ Accepts PDF and image files for text extraction.
✅ Uses **OCR (Optical Character Recognition)** and **regex-based extraction** to convert documents into structured JSON.

2. Missing Data Detection

✅ Identifies missing or incorrectly extracted fields (e.g., invoice number, due date).
✅ Flags incomplete or suspiciously formatted data for review.

3. AI Chat Support for Data Completion

✅ **Interactive chatbot** allows users to confirm missing data.
✅ Uses **GPT-4 and LangChain** to intelligently predict missing values based on extracted document context.
✅ Enables **manual corrections and AI suggestions** through the chat interface.

4. Data Storage & Management

✅ Stores **raw documents, extracted data, and chat interactions** in **Firebase Firestore and Storage**.
✅ Ensures a **secure and scalable** database structure for easy retrieval.

5. Performance Analytics & Insights

✅ Tracks **common missing fields and errors** for continuous model improvement.
✅ Uses **PostHog for analytics** to measure AI accuracy and user interactions.

This AI-powered solution will **increase document processing efficiency, reduce manual intervention, and improve data accuracy**, making it a valuable tool for businesses handling large volumes of documents. 🚀