

AI based PDF DATA READER with ANSWER & QUETIONING MODULE

1. System Overview

A system that:

1. Uploads PDF documents
2. Extracts text, tables, images (optional)
3. Processes the text using NLP
4. Summarizes content into structured notes
5. Enables question-answering (QA) over the document
6. Supports multilingual input + output
7. Provides a clean UI dashboard for searching, summarizing, and QA

2. Full Tech Stack (Recommended Options)

A. Frontend

Category	Technologies
Framework	React.js, Next.js, or Vue.js
UI libraries	Material UI, TailwindCSS, Bootstrap
State management	Redux, Zustand, or Recoil
PDF Viewer	pdf.js, React-PDF
Authentication UI	Firebase Auth UI, Auth0 SDK, or custom JWT UI

B. Backend

Layer	Options
Backend Framework	Node.js (Express / NestJS) or Python (FastAPI / Django)
PDF Text Extraction	PyPDF2, PDFPlumber, Tika, pdf.js server
NLP / AI Models	
→ Text Embeddings	OpenAI Embeddings, HuggingFace models, Google Gemini Embeddings
→ Summarization Models	OpenAI GPT-4/5, Llama 3, BART, T5, Mixtral
→ Question Answering	Retrieval-Augmented Generation (RAG) pipeline using LangChain / Llamaindex
→ Translation / Multilingual	NLLB, M2M100, GPT-4/5, Gemini 2.0 multilingual
Authentication	JWT, OAuth, Firebase Auth

C. Database

Purpose	Recommended Options
User Data / System Config	PostgreSQL, MySQL, MongoDB
Document Storage	AWS S3, Google Cloud Storage, Azure Blob Storage
Vector Database (Required for QA)	Pinecone, Weaviate, Milvus, Qdrant, FAISS

D. Hosting & Infrastructure

Layer	Suggested Platforms
Frontend Hosting	Vercel, Netlify, AWS Amplify
Backend Hosting	AWS EC2 / Lambda, Google Cloud Run, Azure App Service, Railway, Render
Database Hosting	AWS RDS, GCP SQL, MongoDB Atlas
Vector DB hosting	Pinecone Cloud, Weaviate Cloud, Qdrant Cloud
Storage for PDFs	AWS S3, GCP Storage
Authentication	Firebase, AWS Cognito, Auth0

3. System Architecture (High-Level)

1. User uploads PDF → Stored in S3 / Cloud Storage
2. Backend extracts text → **PDFPlumber / Tika**
3. Text split into chunks → **LangChain**
4. Chunks embedded → Stored in **Vector DB**
5. User asks a question →
 - o Query → Embeddings
 - o Fetch top relevant chunks
 - o Pass to LLM (GPT-4/5, Llama3, Gemini)
 - o LLM generates answer
6. Summarization module generates **structured notes** in preferred language.
7. Frontend displays highlights, answers, and downloadable notes.

4. Key Features

A. PDF Extraction Module

- Extract text, tables, metadata, images (optional)
- Clean & preprocess text

B. NLP Processing

- Summarization
- Translation
- Named entity recognition
- Topic modelling
- Concept extraction

C. Question-Answering (QA)

- Document-aware QA using RAG
- Multi-document search
- Answer confidence score
- Citations + source page numbers

D. Multilingual Support

- Input PDF in any language
- Output summary in user-chosen language
- NLP models automatically identify language

E. UI Dashboard

- PDF Viewer
- Search & highlight inside PDF
- Notes download (PDF/Docx)
- History & saved documents

5. Benefits of the System

Business Benefits

- Saves hours of manual reading
- Reduces cost of manual data entry
- Helps teams extract insights faster
- Multilingual flexibility expands global usability

Technical Benefits

- Modular architecture (PDF → NLP → RAG)
- Easy to scale with cloud
- Add more AI features in future
- Secure storage & access management

User Benefits

- Instant summaries
- Ask any question from the PDF
- Accurate answers supported by retrieved text
- Easy interface

6. Future Expansion / Scalability

Add-On Modules

- OCR (for scanned PDFs) using **Tesseract**, **Google Vision**, **AWS Textract**
- Audio summaries (text-to-speech)
- Auto-tagging & classification
- Knowledge graph creation
- Multi-user collaboration
- Teams & roles
- API for external integrations

Enterprise-Level Enhancements

- On-premise LLM deployment
- Data encryption & compliance (GDPR, HIPAA)
- Admin analytics dashboard

7. Estimated Timeline (Realistic Project Plan)

Phase 1 — Requirements & Architecture (1 week)

- Project plan
- Architecture diagram
- Tech stack finalization

Phase 2 — Backend PDF & NLP Pipeline (2-3 weeks)

- PDF extraction
- Text preprocessing
- Embedding generation
- Vector DB integration

Phase 3 — Summarization + Multilingual Module (1-2 weeks)

- Summary templates
- Translation integration

Phase 4 — Question-Answering System (2-3 weeks)

- RAG pipeline
- Retrieval tuning
- Evaluation + testing

Phase 5 — Frontend Development (2-3 weeks)

- Dashboard UI
- PDF viewer
- Upload + result display

Phase 6 — Deployment & Hosting (1 week)

- Backend deployment
- Frontend hosting
- CI/CD integration

Total Estimated Time: 8-12 weeks

1. Tech Stack (Recommended)

Frontend

- **React.js (18+)**
- **React Router v6**
- **TailwindCSS / Material UI / Bootstrap** (choose one)
- **Framer Motion** (for animations)
- **Axios / fetch** for API calls
- **React Query** (optional for data fetching)

Backend (optional for dynamic content)

- Node.js (Express) / Django / FastAPI — depending on your ecosystem
- CMS option: **Strapi** / **Sanity** / **Headless WordPress** (if non-technical team updates content)

2. Website Menu Structure (Dynamic Navigation)

Main Navigation

1. Home
2. Solutions – Products & Services
 - **Products**
 - Artificial Intelligence
 - AI based PDF Data Reader
 - AI based WhatsApp Chatbot (AAIEBOT)
 - AI based WhatsApp Messaging Platform
 - e-Governance
 - Workflow Management
 - MoM Management
 - e-Tapal
 - Visitor Management
 - Administrative Dashboard
 - MIS
 - ERP Solutions
 - School ERP
 - Hostel ERP
 - e-Library
 - LMS
 - Blockchain Technology
 - Decentralised Verifiable Credentials
 - Identity Management
 - App Development
 - React Native Apps
 - **Services**
 - SAP ERP Consulting

- e-Commerce Solutions
 - Cyber Security
 - App Development
 - Tailored Blockchain Applications
 - Tailored AI + Blockchain Applications
 - Full Stack Development
3. Latest Development
 4. Trusted by Global Enterprises & Government
 5. Our Impact (Use Cases)
 6. Strategic Partners
 7. News & Media
 8. Blog
 9. Get in Touch

3. Home Page Layout (React Components)

Sections

- Hero Banner (Company tagline + CTA)
- Solutions Overview (grid cards)
- Products Overview (dynamic slider/cards)
- Services Overview
- Latest Developments
- Partners logo carousel
- Testimonials (optional)
- Footer (company info + links)

4. Dynamic Backend

If you want the content to be **editable**, use:

Best Option: Strapi Headless CMS

- Admin panel to update content
- React fetches content via REST or GraphQL
- Deployed on Render / Railway / Vercel

Dynamic endpoints:

- /products
- /services
- /solutions
- /news
- /blogs
- /partners

5. Optional Enhancements

- Dark/Light mode toggle
- Search across products/services
- WhatsApp contact integration
- Newsletter subscription
- Chatbot (AAIEBOT demo integration)
- Animated counters for “Our Impact”
- Partner logo auto-scrolling

✓ Deliverables I Can Provide (if you want):

- ☞ Full React source code (zip)
- ☞ UI/UX design (Figma layout)
- ☞ Deployment guide
- ☞ Backend (Strapi / Node.js) setup
- ☞ Dynamic API integration
- ☞ SEO optimized page content

1. Recommended Tech Stack (PHP-Based)

Frontend

- HTML5, CSS3, JavaScript
- Bootstrap 5 / TailwindCSS
- jQuery (optional)

Backend

- **PHP 8+**
- **Laravel 10** (recommended for modern web apps)
or
- **Core PHP + MySQL** (if you need lightweight & simple)

Database

- MySQL / MariaDB

Hosting

- **cPanel shared hosting, VPS, or AWS EC2**
- PHP 8.x + MySQL support
- Apache or Nginx server

2. Website Menu Structure (Dynamic Navigation)

Main Navigation

1. Home
2. Solutions – Products & Services
 - Products
 - Artificial Intelligence
 - AI based PDF Data Reader
 - AI based WhatsApp Chatbot (AAIEBOT)
 - AI based WhatsApp Messaging Platform
 - e-Governance
 - Workflow Management
 - MoM Management
 - e-Tapal
 - Visitor Management
 - Administrative Dashboard
 - MIS
 - ERP Solutions
 - School ERP
 - Hostel ERP
 - e-Library

- LMS
 - Blockchain Technology
 - Decentralised Verifiable Credentials
 - Identity Management
 - App Development
 - React Native Apps
 - Services
 - SAP ERP Consulting
 - e-Commerce Solutions
 - Cyber Security
 - App Development
 - Tailored Blockchain Applications
 - Tailored AI + Blockchain Applications
 - Full Stack Development
3. Latest Development
4. Trusted by Global Enterprises & Government
5. Our Impact (Use Cases)
6. Strategic Partners
7. News & Media
8. Blog
9. Get in Touch

Dynamic content loaded from MySQL tables.

3. Admin Panel (PHP Dynamic CMS)

Admin Features

- Login system (username/password)
- Add / Edit / Delete:
 - Products
 - Services
 - Blog Posts
 - Latest Development
 - Partners
 - Contact messages viewer

4. Home Page Layout

Dynamic sections:

- Hero banner
- Solutions overview
- Product categories
- Services list
- Latest development highlights
- Partner logos
- Footer

5. Complete Product & Services List (Ready for DB Insert)

PRODUCTS

AI (Artificial Intelligence)

- AI based PDF Data Reader
- AI based Whatsapp Chatbot (AAIEBOT)
- AI based Whatsapp Messaging Platform

e-Governance

- Workflow Management
- Minutes of Meeting (MoM) Management
- e-Topal
- Visitor Management
- Administrative Dashboard
- MIS

ERP's

- School Management ERP
- Hostel Management ERP
- e-Library
- LMS

Blockchain Technology

- Decentralised Verifiable Document Credentials
- Identity Management

App Development

- React Native Apps for Android & iOS

SERVICES

- SAP ERP Consulting
- e-Commerce Solutions
- Cyber Security Solutions
- App Development
- Tailored Blockchain Applications
- Tailored Blockchain AI Applications
- Full Stack Software Development

6. Optional Enhancements

- SEO optimized pages
- News & media dynamic section
- Multi-language support
- AJAX-based dynamic search
- Testimonials module
- API integration (WhatsApp Bot demo)
- Email automation for contact page

7. I Can Provide Next (choose one):

- ✓ Full PHP project source code (ZIP)
- ✓ SQL dump for all tables
- ✓ Admin Panel (PHP) code
- ✓ Website front-end design (HTML/CSS/Bootstrap)
- ✓ Laravel version of this website
- ✓ Deployment guide for cPanel