
Basic Flow of a PDF Chat SaaS:

User uploads a PDF

You extract the content (text)

You chunk the text (break into pieces)

You convert those into embeddings (vectors)

When user chats, you:

1. Search similar chunks
2. Pass the best chunks + user's question to ChatGPT

Estimated Cost Per User (GPT-3.5 vs GPT-4)

Assume a **single user uploads a 30-page PDF** and asks **20 questions**.

Step 1: PDF Parsing (Free or cheap)

Use open-source tools (like `pdf-parse`, `PyMuPDF`, or `pdfplumber`)

Cost: ~\$0 (local or server-side script)

Step 2: Text Embedding Cost

You use OpenAI's `text-embedding-3-small`:

Cost: \$0.00002 per 1K tokens

30 pages \approx 10K tokens

Cost: **\$0.0002** (super cheap!)

Step 3: Chat With GPT (Main cost)

Each question = you send ~2K tokens (context + question)
GPT-3.5 Turbo:

Input: $2K \times \$0.0015 = \0.003

Output: $500 \times \$0.002 = \0.001

Per query: ~**\$0.004** 20 questions = **\$0.08 (~₹6.6)**

GPT-4 Turbo:

Input: $2K \times \$0.01 = \0.02

Output: $500 \times \$0.03 = \0.015

Per query: ~**\$0.035** 20 questions = **\$0.70 (~₹58)**

Final Cost Estimate (Per User, 1 PDF, 20 Questions)

Model	Embedding	Chat (20 Qs)	Total Cost
GPT-3.5 Turbo	~\$0.0002	~\$0.08	~ \$0.09 (₹7.5)
GPT-4 Turbo	~\$0.0002	~\$0.70	~ \$0.71 (₹59)

Optional Infra Costs (Per Month Estimate)

Resource	Free Option?	Paid (Est.)
File Storage (PDFs)	Firebase / Supabase free	\$5 - 10/month
Vector DB (Embeddings)	Free: SQLite, Qdrant	Paid: Pinecone ~\$10/month
Backend Hosting	Free: Vercel, Render	\$5 - 20
DB (Users, uploads)	Supabase free tier	\$0 - 10
Domain & SSL	₹500/year	—

Monetization Strategy Suggestion

Plan	Monthly Fee	Max Upload	GPT Model
Free Tier	₹0	1 PDF / 5 Qs	GPT-3.5
Pro Plan	₹199/month	10 PDFs / 100 Qs	GPT-3.5
Premium	₹499/month	Unlimited	GPT-4 Turbo

How Your Credit System Will Work

User gets some free credits (e.g. 20 questions)

You charge small amounts like:

₹29 → 100 credits

₹99 → 500 credits

₹199 → 1500 credits

Each **chat with a PDF** costs **1 credit**

OR based on token usage: 1 credit = 500 tokens

If user has **0 credits** → only Free Plan active

Credits stored in database (e.g. Supabase, Firebase, PostgreSQL)

Sample Credit Pricing (India-style)

Pack Price	Credits	Cost per Chat	Notes
Free ₹0	10 credits	~10 chats	Once, per user
Mini ₹29	100 credits	₹0.29/chat	Great for testing
Lite ₹99	500 credits	₹0.19/chat	Popular plan
Pro ₹199	1500 credits	₹0.13/chat	Power users

Credit Usage Example

Each message (chat with PDF):

GPT-3.5 = 1 credit

GPT-4 Turbo = 3 credits

How to Build This (Tech Stack Suggestion)

Feature	Stack Suggestion
PDF Upload + Extract	Node.js + pdf-parse or Python
Embedding	OpenAI Embedding API (text-embedding-3-small)
Vector Search	Qdrant (free & fast) or Supabase vector extension
Chat AI	GPT-3.5 / GPT-4 via OpenAI API
Credit System	Supabase DB: users, credits, usage_log
Frontend	React + Tailwind (smooth UI)
Auth	Supabase Auth (email login / Google login)
Payment	Razorpay or Stripe (for credit top-up)
Hosting	Vercel or Render (free for small projects)

Example Database Structure (Supabase)

users

id email plan credits created_at

usage_log | id | user_id | type (chat/pdf) | tokens_used | credits_used | timestamp |

payments | id | user_id | amount | credits_added | payment_status | date |

Bonus Tip:

Add a **"Watch Ad for 1 Credit"** option later to earn passive income via ads too
