

# Internship Assignment – PDF Viewer + Data Extraction Dashboard

## Objective

Build a PDF Review Dashboard where users can: - Upload and view PDFs in-browser - Run AI extraction using Gemini API or Groq - Edit extracted data, perform CRUD, and store in MongoDB - Keep frontend and backend separated in a monorepo - Follow basic best practices (clear structure, linting, error handling)

## Tech (must use)

- Monorepo (Turborepo or npm workspaces)
- apps/web    Next.js (App Router) + TypeScript + shadcn/ui
- apps/api    Node.js (TypeScript)
- Database: MongoDB (Atlas)
- AI: Gemini API or Groq
- PDF viewer: pdf.js (UX reference: tomoki1207/vscode-pdfviewer)
- Deploy: Both web and API on Vercel

## Core Features

- PDF Viewer: Upload PDF ( ≤ 25 MB), render in left panel, store file (Vercel Blob or MongoDB GridFS)
- AI Data Extraction: Extract key invoice data (vendor, invoice, line items) using Gemini or Groq
- Data Editing & CRUD: Edit fields, perform Create/Read/Update/Delete in MongoDB, searchable list view
- API (apps/api): REST JSON endpoints for upload, extract, invoices CRUD
- UI (apps/web): Split layout (Left = PDF viewer, Right = editable form), buttons for Extract, Save, Delete

## Minimal Data Shape (MongoDB)

```
{ fileId: string, fileName: string, vendor: { name: string, address?: string, taxId?: string }, invoice: {
  number: string, date: string, currency?: string, subtotal?: number, taxPercent?: number, total?:
  number, poNumber?: string, poDate?: string, lineItems: Array<{ description: string, unitPrice:
  number, quantity: number, total: number }> }, createdAt: string, updatedAt?: string }
```

## Repo & Deployables (deliverables)

- Public GitHub monorepo with apps/web and apps/api
- Deployed links (Vercel)
- 3–5 min demo video: upload    view    extract    edit    save    list    update    delete
- README.md with setup, env vars, local run, deployed URLs, API docs

## Acceptance Criteria (pass/fail)

- PDF uploads and renders in viewer
- AI extraction works (Gemini or Groq) and returns valid JSON
- Edited fields persist in MongoDB; list view with search
- All API endpoints respond correctly
- Both apps run locally and are live on Vercel