

Paggo - OCR Case Outline (5 days)

Objective:

Design and implement a solution that allows users to upload documents to a webpage, get the extracted text (OCR) and request interactive explanations over the extracted data (LLM).

Requirements:

Database:

- Model the database using Prisma ORM

Frontend:

- Develop a simple webpage using React (we recommend the Next.js framework) that allows users to upload an invoice image.
- Implement feedback mechanisms on the webpage (e.g., progress indicator, success/error messages).

Backend:

- Create a backend service using the NestJS framework to handle image uploads, manage OCR processing, LLM integration and store results in the database.

Authentication and Permissions:

- Users must be authenticated prior to being able to upload the invoice. You can use any open-source authentication tool and decide if user auth will be handled by the Next app or the Nest API.

System Features:

- **LLM Integration for Context and Explanation:** Integrate a Language Model (e.g., GPT-4) to explain or provide context for the extracted text to the users. Enable user queries about extracted information.
- **View Uploaded Documents:** Allow users to see a list of all their previously uploaded documents along with the extracted information and LLM interactions.
- **Download Documents with Extracted Text:** Provide an option for users to download the uploaded documents with the appended extracted text and LLM interactions.

Deliverables and criteria:

- A link with the repository source code for the frontend and backend services.
- Instructions for setting up and running locally the solution
- A link with the deployed app (we would advise using Vercel for its simplicity)
- We will evaluate the application of best development practices in both the frontend and backend.
- We will favor working solutions, if all features can't be completed, focus on the most critical ones.