

Congratulations on making it this far. Welcome to the final round!

This is a 3-hour project.

In this project, you'll have to develop a production-ready application that takes in shipment documents and extracts data for the user. You're free to use any AI tools (OpenAI, Claude, Gemini, Cursor, GitHub Copilot, etc). No help from other humans is allowed.

Use your best judgment on what to prioritize. You'll have to commit all your code at the end of the 3 hours. You are given a starter code attached at the end of the document

You'll be judged on what was completed, attention to detail, error handling, coding practices, and robustness of the software.

Feel free to be creative, and let us know any assumptions made at the end.

Project Components

API

Create an API that takes in a list of document files and extracts relevant data. There can be multiple documents (PDFs and/or XLSX) related to a single shipment

Documents

For testing, you'll be given two documents:

- A bill of lading (.pdf)
- A commercial invoice and packing list (.xlsx)

LLM

You can use Anthropic API key:

sk-ant-api03-T2GA0yHD0E4kmiEi_n7njFTkz_p_vDjQJBfiyM3ce7ng0PFPxU9l2VgRI6JKvxetx
cY26vxejYKi05qz4sAZAw-rwDuSAAA

UI

Create a platform with the following functionalities:

1. Upload the documents to extract data.
2. show an editable form with prefilled data (fields to be extracted:
<https://forms.gle/11kUya5nTebvFBgn7>)
3. Option to view the documents on the side with the extracted data for easy audit

Bonus Work:

Deployment

Make the application production-ready by containerizing it with Docker. Include all dependencies and setup network between the containers

Evaluation

Come up with an eval script that calculates accuracy, precision, recall, and F1 for the given set of documents. This will help us understand how robust the product is.

Testing

Write unit tests for the functions. You can use pytest library for the unit tests.

Boilerplate

Link: <https://github.com/Amari-AI/project-boilerplate>

You have to fork this boilerplate and at the end of 3 hours, create a pr against the main branch of your forked repo

Be aware that there might be multiple issues in the current boilerplate.

The purpose is to provide you with a starter code. Reuse the existing code as much as possible and change the template to best fit your solution. Don't be afraid to employ creative solutions