Tabularis

Developing a Vision-Language Model (VLM) architecture that can efficiently detect efficiently detect and extract structured tables from documents and also answers answers queries based on them

CREDITS

TABULARIS

Meet the team

Akshit Bhola

Kaustubh Srivastava

Saurabh Singh

Mohit Kumar

Bharat Agrawal

Vivek Kumar

GROUP-22

Use case

Finance teams receive thousands of invoices in varying formats — PDF scans, digital forms, and images — each with different table layouts for line items, taxes, and totals. Manual data entry is time-consuming, error-prone, and expensive.

Impact:

- Reduction in manual data entry time
- Improved accuracy over rule-based OCR tools
- Scalable across vendors and document formats
- Natural Language Querying across tables

Our Goals

1. Accurately detect and localize tables in diverse document formats
2. Extract the layout and content of tables, including rows, columns, and cells
3. Combine visual layout and textual content to build a structured representation (CSV/HTML)
4. Understand and answer natural language queries based on table data

TABULARIS

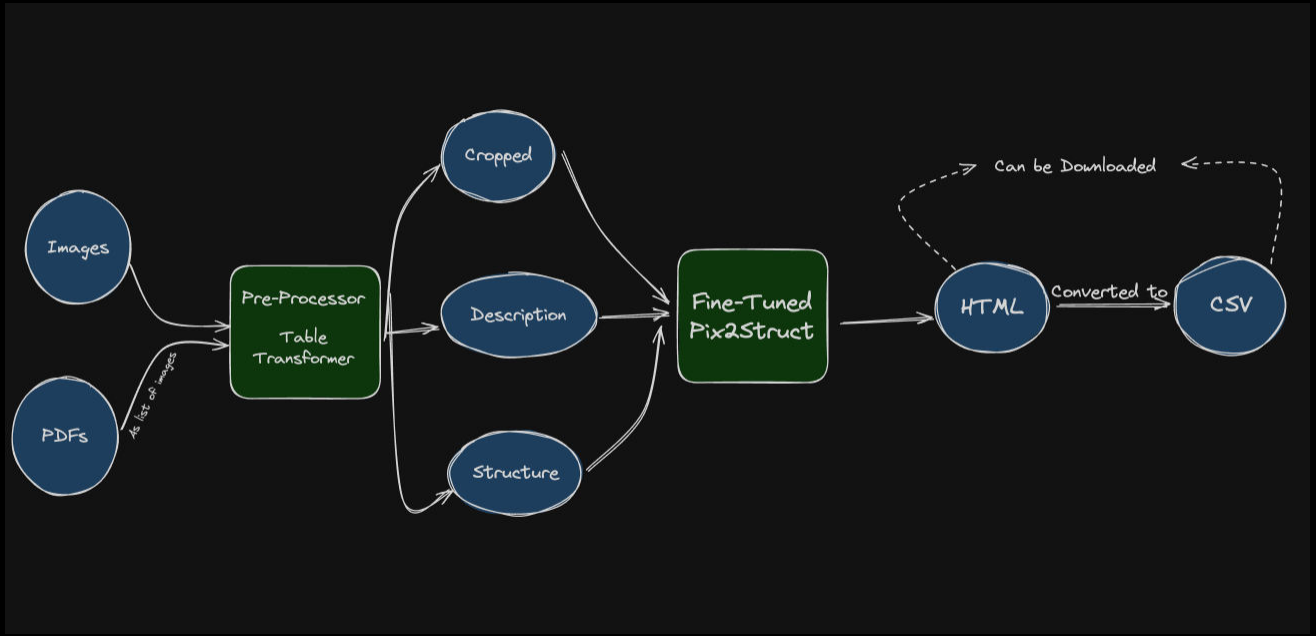
Our Journey

GROUP-22

OVERVIEW

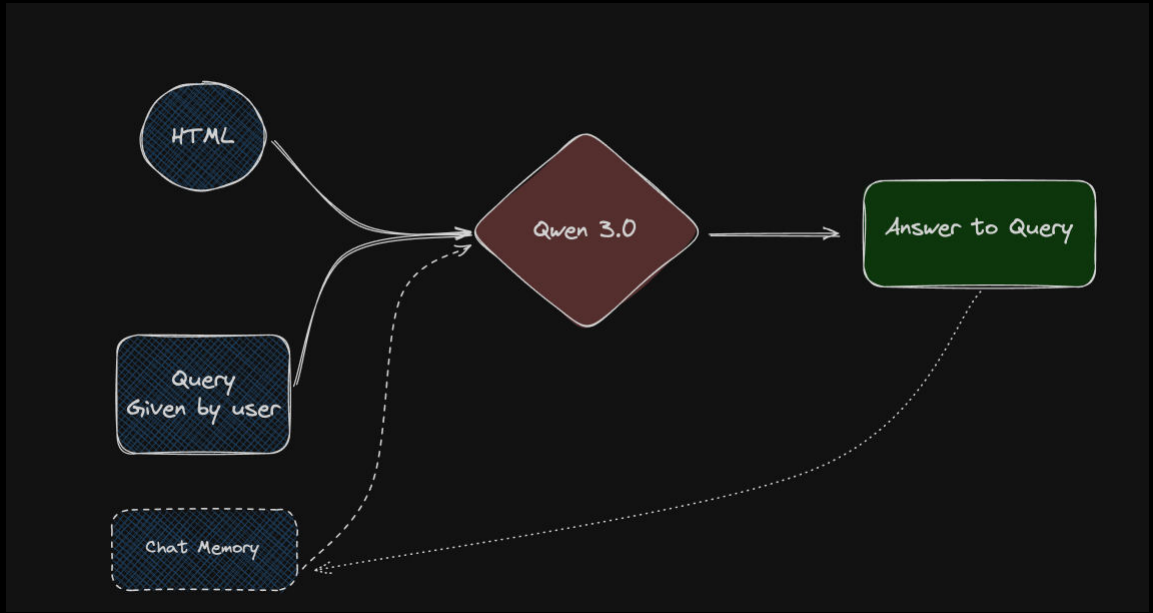
TABULARIS

Image to HTML Architecture



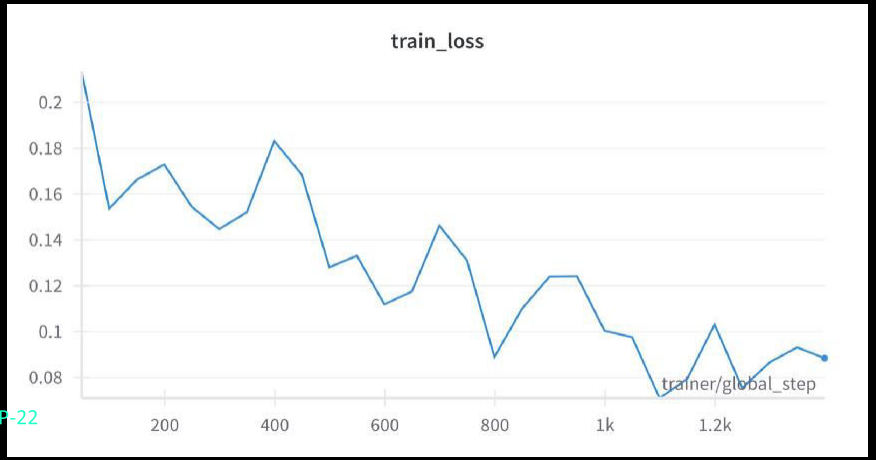
GROUP-22

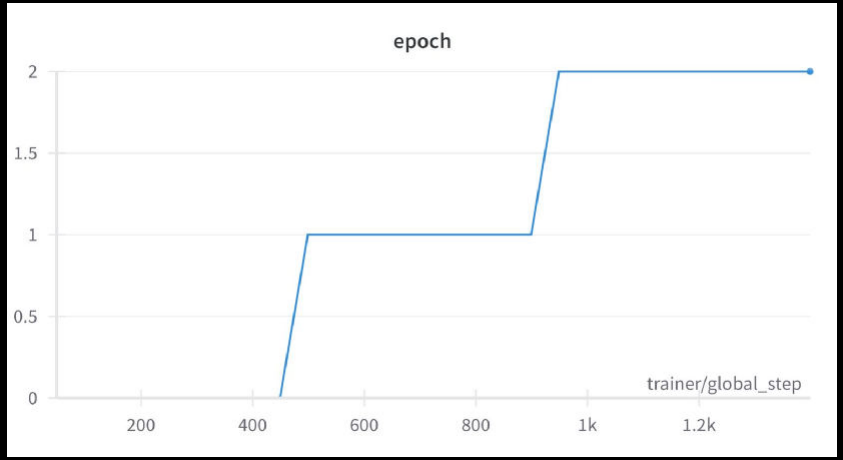
Query-Processing Architecture



TABULARIS

Fine-Tuning Pix2Struct





TABULARIS

Validation metrics

MODEL	PRECISION	RECALL	GRITS
TABLE TRANSFORMER	0.902	0.9305	0.9849

Technologies Used

- Python
- Wandb.ai (Fine Tuning)
- Detection Transformer
- PyTorch
- NextJS
- Flask
- Tailwind