Automated Medical Data Entry Using LLMs

Anju Santhosh Kumar, Yajie Li, Robin Dziewietin, and Virginia Partridge

Motivation

- **Doctors Without Borders** provides emergency aid around the globe in war, disease, and disaster stricken areas.

- Currently, **aggregate clinic data** required for data analytics is collected through **paper forms** and manually entered into an **online database**.

- **Manual data entry is slow**, leading to reporting delays and human errors.

- Allow staff to spend their valuable time providing medical care, not data entry
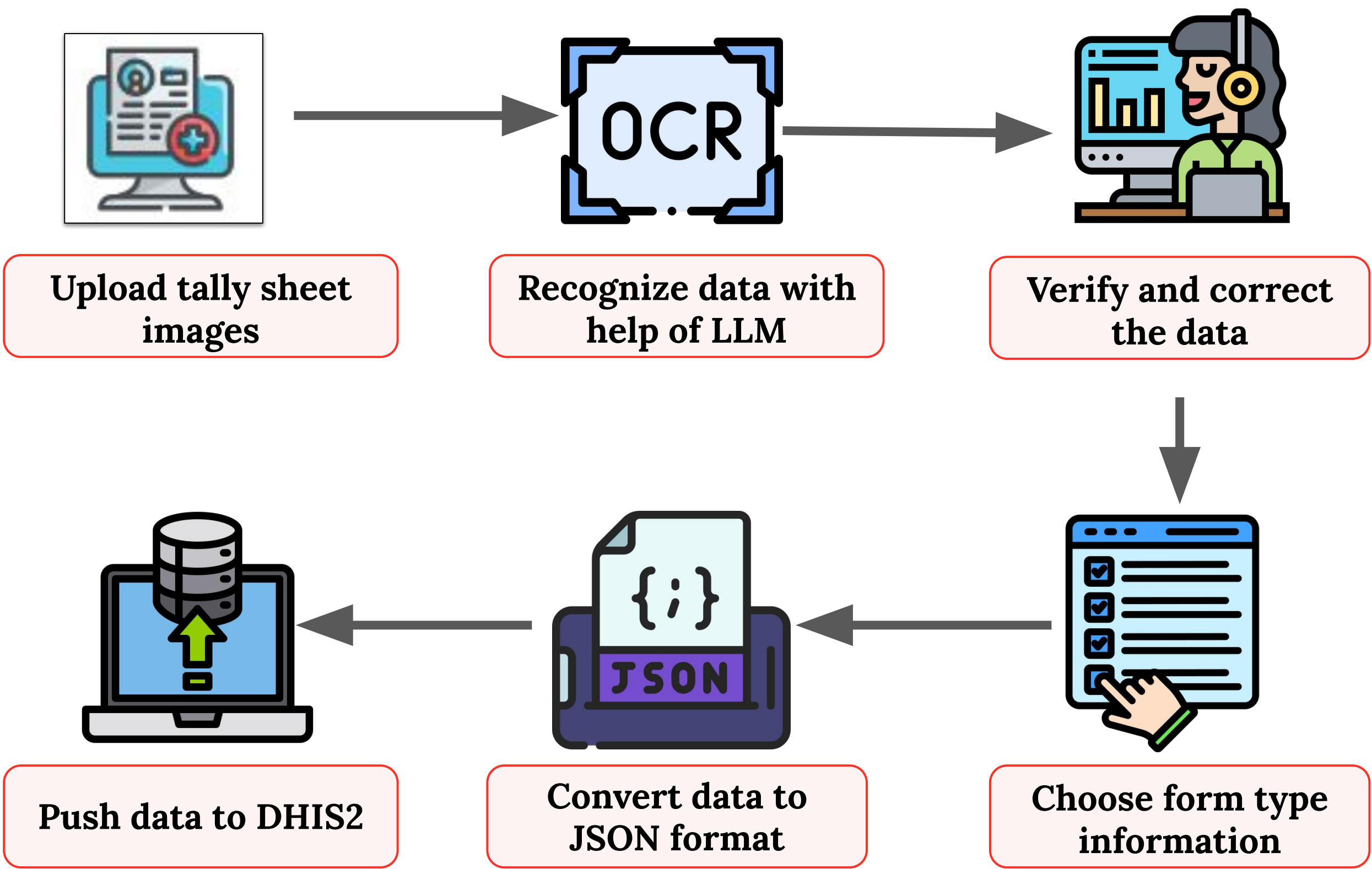
    Goal : How can we speed up data processing and reduce human errors?

Summary

- An interactive Streamlit web application that allows users to upload images, verify and edit AI-recognized content and upload to the DHIS2 health information management system.

- Features:
  - User authentication
  - Batch process multiple images
  - Page-wise view of images
  - Human-in-loop verification
  - Progress tracking for verification
  - One-click data upload to DHIS2

Pipeline

Upload tally sheet images	Recognize data with help of LLM	Verify and correct the data
Push data to DHIS2	Convert data to JSON format	Choose form type information

Software architecture

dhis2 → Auth → Streamlit ← Payload
Streamlit → Prompt → Azure / OpenAI ← Response

Results

Input — Output

Doctors Without Borders Image Recognition Data Entry

Clear Form

Page Number
1 ✓

Show Image

ANC - diagnostics

	0	1	2	
		Pop1: Resident	Pop2: Displaced	Add Column
First ANC visit this pregnancy		65	05	Column to delete
Follow-up ANC visit		178	08	0
Syphilis number tested		65	05	Delete Column
Syphilis number positive				
HIV number tested (first test in ANC)				
HIV number positive (first test in ANC)				
HIV number tested (later visit)				
HIV number positive (later visit)				
Hep B number tested		66	05	

Scan to checkout code base!

Evaluation

- Testing on the Doctors Without Borders data set:
  - **GPT-4o**: >90% accuracy.
  - **Traditional OCR tools**: ~85% accuracy with significant variation.

- **Positive initial feedback**:
  - **Time**: less time spent entering data.
  - **Errors**: GPT-4o's accuracy makes errors easier to spot.

Future

- Further testing and development may lead to **full deployment**.

- **Image preprocessing**:
  - More reliable image recognition.
  - Inform users about best practices.

- **Indicators** for low-confidence entries **to direct focus**.

Partner

Médecins Sans Frontières (MSF), translated as Doctors without Borders, is an international, independent medical humanitarian organisation.

MSF provides medical assistance to people affected by conflict, epidemics, disasters, or exclusion from healthcare. MSF's actions are guided by medical ethics and the principles of independence and impartiality.