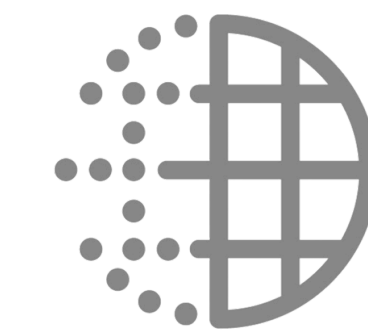


Diagnolingo

Kushagra Agarwal

Mentor: Eric Nyberg (Language Technologies Institute)



Carnegie Mellon University
Master of
Computational
Data Science

Introduction

In Indian outpatient clinics, hand-written prescriptions create fragmented medical histories, hindering effective patient care. The traditional healthcare approach lacks an integrated Electronic Health Record (EHR) system, creating a barrier to consistent medical documentation.

Diagnolingo proposes an innovative AI-powered EHR system designed specifically for India's unique healthcare environment. It aims to capture doctor-patient conversations in multiple Indian languages, convert these into structured EHR entries, and utilize this data to provide predictive insights for better healthcare outcomes.

Project Goals & Hypothesis

This project aims to **automate medical scribing in India's outpatient clinics and integrate with EHR systems** under the ABHA program. With advancing digital healthcare initiatives, regulatory support, and mature AI technologies, its implementation is well-timed.

We expect to:

- Achieve 95% accuracy in transcribing Indian languages using OpenAI Whisper, evaluated via Word Error Rate (WER).
- Ensure 95% top-3 accuracy in assigning ICD, NDC, and CPT codes using Retrieval Augmented Generation (RAG).
- Improve clinical predictions using EHR data, outperforming baseline models (future scope).
- Create an India-specific EHR schema, validated by providers.
- Reduce clinician workload and enhance patient outcomes, measured by time saved, diagnostic accuracy, and satisfaction.

Related Work

Existing solutions like Augmedix [1] and Abridge [2] automate clinical note-taking but focus primarily on English-speaking contexts. Diagnolingo is uniquely tailored to India's healthcare challenges, offering:

- Vernacular Language Processing:** Handles India's linguistic diversity, including code-switching between English and regional languages during conversations.
- Automated Medical Code Generation:** Automates SOAP note annotation with ICD, NDC, and CPT codes using Retrieval Augmented Generation (RAG), unlike solutions relying on manual input.
- Cost-Effectiveness:** Employs open-source and affordable models, addressing economic constraints in Indian healthcare, unlike competitors using costly APIs like Google's MedPalm2.

Requirements

1 Intended Users

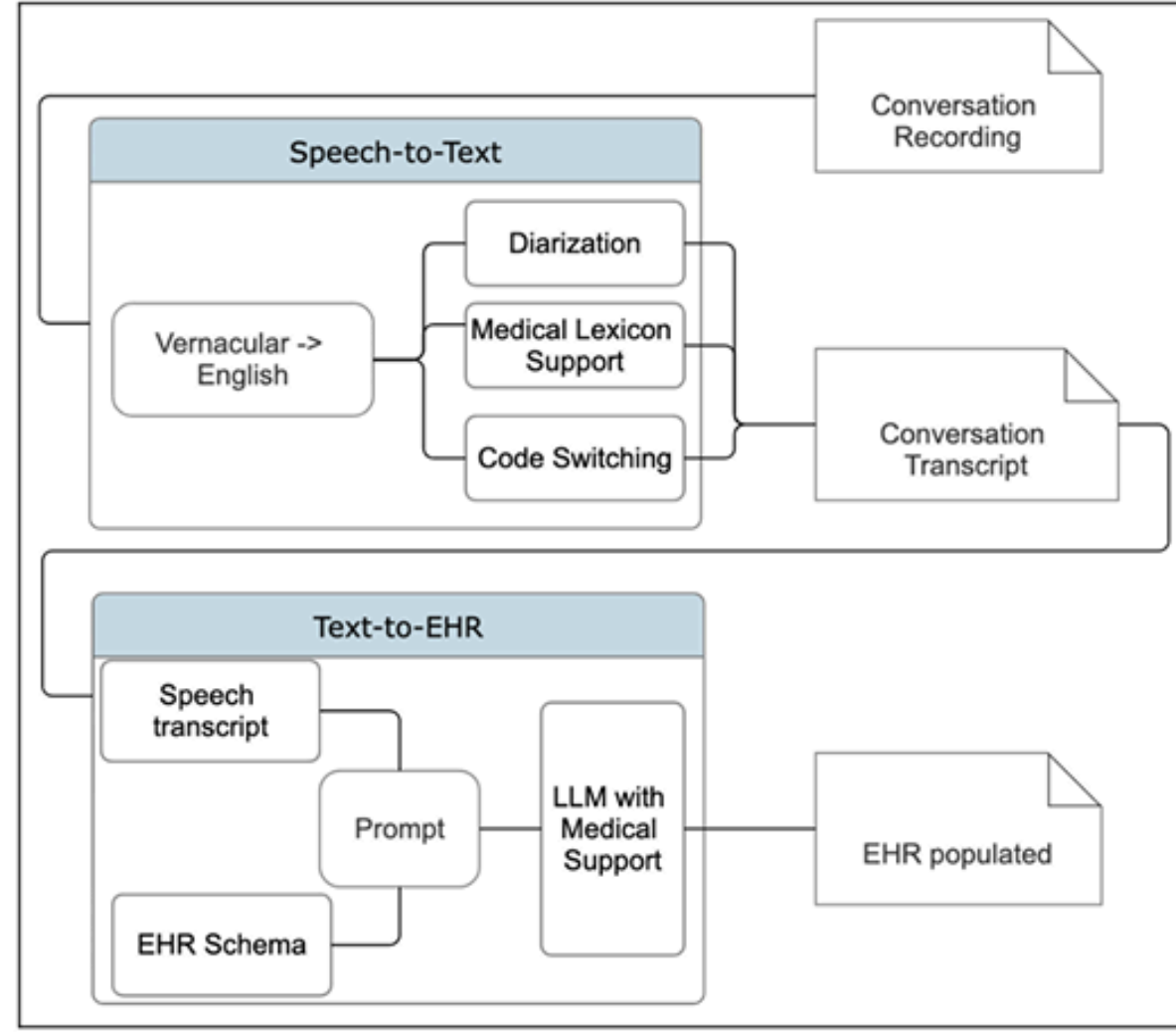
- Doctors:** Primary users for documenting patient interactions, reviewing transcriptions, and accessing EHRs.
- Patients:** Secondary users for viewing medical records, managing appointments, and engaging with healthcare providers.

2 Feature Requirements: Doctors will need voice recording, transcription review, EHR integration, and predictive analytics tools, and patients will require a simple interface for medical history.

3 System Requirements

- Functional:** Accurate transcription & seamless EHR mapping.
- Non-Functional:** High accuracy, performance, scalability, and security to meet healthcare data standards.
- Resource:** Data for validation, hardware for system operations, and human resources for development and maintenance.

System Design Overview

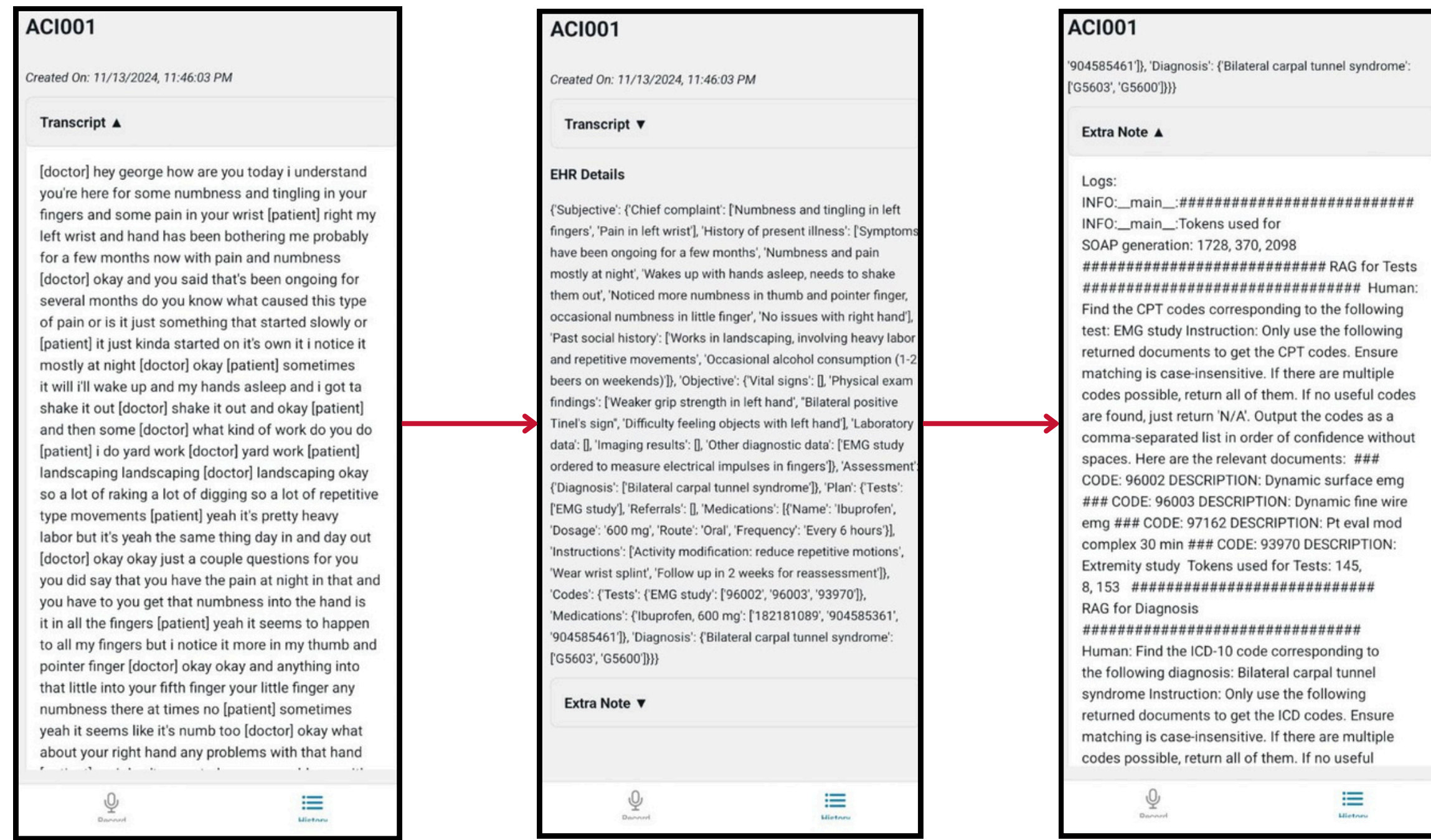
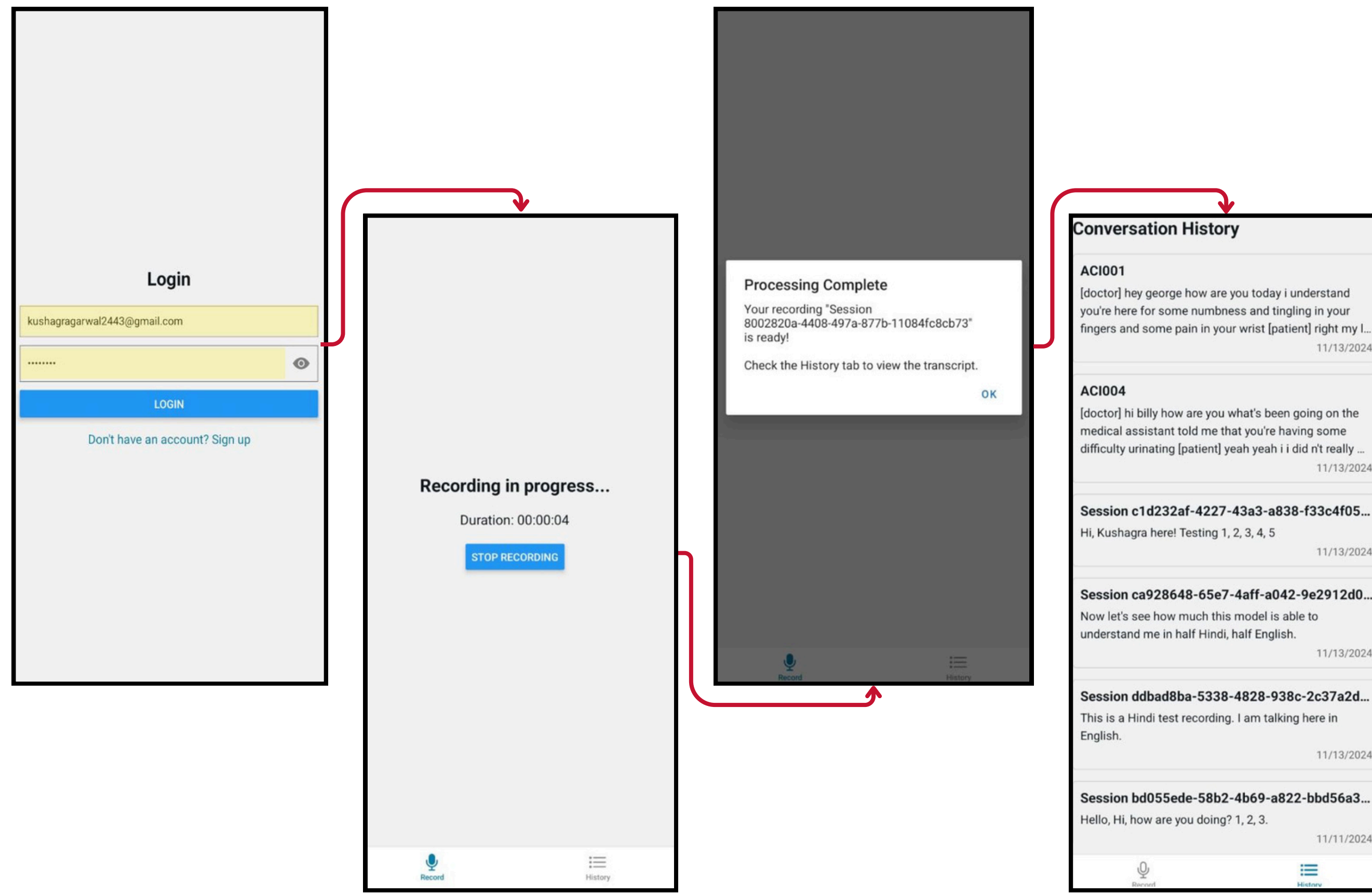


The layered architecture includes:

- Speech-to-Text Module:** Converts doctor-patient conversations into English text, supporting regional languages.
- EHR Population Module:** Utilizes large language models (LLMs) to organize patient information, generate ICD, NDC, and CPT codes, and populate the EHR database.
- Predictive Analytics Engine:** Delivers diagnostic insights.

The workflow begins with conversation recording, followed by transcription, EHR generation, provider review, and analytics application. Key components interact seamlessly, supported by robust backend infrastructure for data management and secure access mechanisms.

Diagnolingo Mobile App



Results

1) Speech-to-Text

Using Whisper, we evaluated transcription quality for Hindi audio samples with word error rate (WER), character error rate (CER), unigram and bigram similarities, and medical concept metrics (precision, recall, F1).

- Non-Medical Metrics:** Average **WER: 0.07, CER: 0.002**, unigram similarity: 0.31, bigram similarity: 0.11.
- Medical Metrics:** **Recall: 0.854, Precision: 0.903, F1 score: 0.818.**

2) Text-to-EHR

The transcript-to-SOAP note conversion used a detailed prompt for structured EHR generation and Retrieval-Augmented Generation (RAG) for accurate ICD, NDC, and CPT codes.

- Challenges:** Some CPT codes were missing due to non-matching descriptions.
- Physician Feedback:** Initial EHR samples scored a high **3.8/5**
- Metrics:** **Precision: 0.526, Recall: 0.533**, and **F1 score: 0.527**, are promising for subsequent iterations after incorporating doctor feedback.

Error Analysis

We evaluated the Text-to-EHR pipeline for ICD, NDC, and CPT codes:

- ICD Codes:** For “Bilateral carpal tunnel syndrome”, RAG prioritized G5603, correctly identifying the bilateral condition.
- NDC Codes:** Multiple codes for “Ibuprofen, 600mg” highlighted manufacturer-specific redundancy. Consolidating codes by active ingredients could simplify selection.
- CPT Codes:** For “Order EMG for fingers”, RAG discarded dynamic codes, identifying the test as static, showcasing contextual understanding.

While RAG performs well contextually, addressing NDC redundancy and CPT limitations can enhance clinical integration.

Future Work & Conclusions

Post-capstone, Diagnolingo will evolve from prototype to market-ready, focusing on predictive analytics, rigorous testing, and iterative improvements based on doctor feedback. The next steps include deploying the prototype to users, refining features through real-world feedback, and building a data-driven foundation for sustained innovation and market growth.

Acknowledgements

I would like to thank my brother, Akshat Agarwal, for his support and creative ideas while brainstorming through technical difficulties.

References

- <https://www.augmedix.com/>
- <https://www.abridge.com/>

