
Diagnolingo

Kushagra Agarwal and Eric Nyberg
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213
kagarwa2@andrew.cmu.edu

Abstract

1 Diagnolingo is aimed to revolutionize outpatient clinic operations in India with an
2 AI-powered Electronic Health Record (EHR) system tailored for small practices.
3 Imagine a system that seamlessly captures doctor-patient conversations in vernacular
4 regional languages, automatically populating EHR entries. Our solution would
5 allow the use of advanced analytics to provide personalized predictive insights,
6 empowering doctors with valuable diagnostic support. By merging speech translation
7 with automated EHR data entry, we're not just streamlining operations but
8 enhancing patient outcomes by reducing mis-diagnoses and missed-diagnoses. All
9 of this will be provided as an application for the doctors and patients to interface
10 with.

11 1 Introduction

12 In outpatient clinics across India, the prevalent practice of maintaining hand-written paper prescriptions
13 results in fragmented medical histories, leading to a lack of standardized longitudinal patient
14 records [7,8]. This scenario hinders the effectiveness of patient care, as every doctor visit essentially
15 starts from a blank slate, disregarding the patient's medical history. The traditional approach in
16 Indian healthcare lacks an integrated Electronic Health Record (EHR) system, which is crucial for
17 consolidating patient data over time. This gap in healthcare delivery is not just an inconvenience but a
18 significant barrier to providing consistent and informed patient care [9]. Globally, there are solutions
19 aimed at digitizing healthcare records and enhancing the efficiency of patient-doctor interactions. For
20 instance, companies like Augmedix in the USA focus on automating clinical notes from doctor-patient
21 conversations. However, these solutions are not fully applicable to the Indian context, primarily due
22 to the unique challenges India faces, including diverse vernacular languages, economic constraints
23 that limit the use of expensive APIs, and the specific needs of Indian outpatient clinics.

24 Diagnolingo proposes an innovative AI-powered EHR system designed specifically for India's unique
25 healthcare environment. It aims to capture doctor-patient conversations in multiple Indian languages,
26 convert these into structured EHR entries, and utilize this data to provide predictive insights for better
27 healthcare outcomes. This solution is distinct in its ability to understand and process vernacular
28 languages, its cost-effectiveness and design that is specifically tailored to the Indian outpatient clinic
29 workflow. With the Indian government keen on integrating EHRs into the healthcare system under
30 the Ayushman Bharat Digital Mission (ABHA) [1,2,3,4], Diagnolingo stands at the forefront of this
31 transformative wave.

2 Hypothesis/Project Goal

The goal of the project is to be able to introduce and automate the process of medical scribing of outpatient clinical conversations in India and also integrate with the upcoming EHR systems that are coming up under the ABHA program mentioned in the introduction. The project is being developed at "just" the right time in the healthcare scene in India, with regulations opening up and the government looking forward to building a digital healthcare ecosystem (under ABHA). The Food and Drug Administration (FDA) has also opened up the AI/ML-powered Software as a Medical Device (SaMD) 510(k) submission pathway, approving more than 1000 companies by now. All of these factors coming together at the right time, along with the maturity of large language models, increase the odds of this proposed solution to work. The empirical and formative hypotheses mentioned below are paramount for this success.

2.1 Empirical Hypotheses

1. **Transcription Accuracy:** We hypothesize that we can transcribe doctor-patient conversations in multiple Indian languages and dialects with a minimum accuracy of 95%. This is predicated on leveraging advanced speech recognition technologies (OpenAI Whisper). The testable metric here would be the Word Error Rate (WER) compared against a benchmark dataset of transcribed medical conversations (namely ACI bench introduced later).
2. **Medical codes accuracy:** The medical codes assigned to the SOAP note (Subjective, Objective, Assessment, Plan) namely, ICD codes for diagnosis, NDC codes for medications, and CPT codes for procedures should be accurate. We retrieve the top 3-4 codes using RAG (Retrieval Augmented Generation) and we would would a top-3 accuracy of atleast 95% for each of the three code types.
3. **Predictive Diagnostic Accuracy:** We assert that utilizing longitudinal EHR data through Diagnolingo will provide predictive insights that are more accurate and clinically relevant than baseline models. This can be evaluated through comparative studies against existing clinical risk assessment tools, employing metrics like sensitivity, specificity, and Positive Predictive Value (PPV). This will happen later in the project and is not in the scope of the Capstone itself.

2.2 Formative Hypotheses

1. **EHR Schema Effectiveness:** We hypothesize that a customized EHR schema for the Indian context will capture essential patient data efficiently and comprehensively. This will be assessed through feedback from healthcare providers, focusing on the schema's usability and the relevance of the data captured.
2. **System Impact on Healthcare Delivery:** We aim to significantly reduce the workload of healthcare professionals while enhancing patient outcomes. This hypothesis will be tested by measuring the time saved in patient data handling, the reduction in diagnostic errors, and improvements in patient satisfaction scores.

3 Relationship to Prior Work

While existing solutions by companies like Augmedix and Abridge have pioneered the automation of clinical notes from doctor-patient interactions, Diagnolingo distinguishes itself by addressing the specific challenges and needs of the Indian healthcare ecosystem. The project's innovation lies in its comprehensive approach, which includes vernacular language processing, a customized EHR schema, and automatic medical code generation.

1. **Language and Code-Switching:** Unlike existing solutions that primarily focus on English, Diagnolingo is designed to handle the linguistic diversity of India, including code-switching between English and regional languages during the same conversation. This capability is

78 crucial in a multilingual country like India, where the dialect and language vary significantly
79 across regions.

80 2. **Medical code generation:** Other solutions in the market still leverage manual annotation
81 by a medical practitioner of the SOAP note with medical codes, however, we automate this
82 process using RAG.

83 3. **Cost-Effectiveness and Open Source Models:** Acknowledging the economic constraints of
84 Indian healthcare, Diagnolingo leverages cheaper models, contrasting with other solutions
85 that rely on costly APIs, such as Google's MedPalm2 use by Augmedix.

86 4 Fall Semester Development Goals

87 Over the Fall semester, I have made significant strides in developing Diagnolingo. I began by
88 gathering and consolidating relevant datasets, specifically ACI Bench and MTS Dialog, which contain
89 doctor-patient conversation transcripts and corresponding EHR notes. These datasets provided a
90 foundation for validating the system. Using the ACI Bench transcripts, I focused on creating SOAP
91 notes through prompt engineering, ensuring that the generated notes were concise, well-formatted,
92 and structured as JSON objects and relevant in the Indian context. This format allows the notes to be
93 directly integrated into the database, streamlining the process and setting a reliable benchmark for
94 EHR note accuracy when compared with existing notes in the ACI Bench dataset.

95 One of the core innovations in Diagnolingo is the automatic generation of medical codes to enrich
96 the EHR notes. To achieve this, I integrated a Retrieval-Augmented Generation (RAG) approach
97 to retrieve the top 3-4 relevant codes for each SOAP note, covering ICD codes for diagnoses, NDC
98 codes for medications, and CPT codes for procedures. I created separate vector databases for each
99 code type to support efficient retrieval, ensuring high accuracy and relevance. This process allowed us
100 to construct comprehensive EHR notes by combining the SOAP notes with automatically generated
101 medical codes, enhancing the clinical value of the notes without manual annotation.

102 We have built the application's front end using React and the backend with FastAPI. The goal is to
103 develop a system that can handle doctor-patient conversations in vernacular languages, translate them
104 into English, and convert the transcripts into structured EHR notes. This involves using OpenAI's
105 Whisper model to transcribe audio conversations from Indian languages into English. We lack
106 real-world Indian-language conversational data, therefore, I have created sample conversations in
107 Hindi by translating English transcripts from the ACI Bench dataset. This mock data allows us to test
108 and refine the transcription and EHR generation pipeline.

109 In summary, my work this semester has focused on assembling the foundational elements of Diag-
110 nolingo, including data acquisition, SOAP note generation through prompt engineering, automated
111 medical coding via RAG, the development of an initial application interface, and rigorous testing.
112 Going forward, I plan to refine the application pipeline, and expand testing to validate the system's
113 accuracy and effectiveness in real-world clinical settings (both qualitative and quantitative).

114 5 Project Requirements

115 5.1 Intended Users

- 116 • Doctors will be the primary users who will leverage the system for documenting patient
117 encounters, accessing medical histories, and utilizing AI-driven diagnostic insights.
- 118 • Patients will be the secondary users who can access their medical records and engage with
119 healthcare providers through the system.

5.2 Differences in Required Features

Doctors require comprehensive functionalities for voice recording, transcription review, EHR data entry, and access to predictive analytics. Patients need a user-friendly interface to view their medical histories, upcoming appointments, and potentially communicate with their healthcare providers.

5.3 System Functionality and Requirements

- Functional Requirements include accurate transcription, efficient EHR data mapping, and predictive analytics for healthcare insights.
- Non-Functional Requirements include focus on accuracy, performance, scalability, and security, ensuring the system is reliable and compliant with healthcare data standards.
- Resource Requirements include encompassing the data to validate the pipeline, hardware for supporting system operations, and human resources for development and maintenance.

6 Experiment/System Design Overview

A lot of emphasis was placed on developing a robust and scalable architecture that could handle diverse languages and dialects, integrate seamlessly with healthcare workflows, and provide reliable predictive analytics. The architecture is layered, starting with the speech-to-text module, followed by the transcript processing layer for EHR mapping, and culminating in the analytics engine for predictive insights. Implementation includes the deployment of fine-tuned AI models, a user-friendly interface for interaction, and backend infrastructure for data management and analysis.

The workflow begins with the doctor-patient conversation recording, followed by transcription, EHR entry generation, review by the healthcare provider, and the application of predictive analytics for diagnostic support. The system architecture of Diagnolingo is structured to facilitate seamless integration of its various components, enabling efficient processing of doctor-patient conversations and generation of Electronic Health Records (EHRs). The following diagram provides an overview of the high-level components and their interactions:

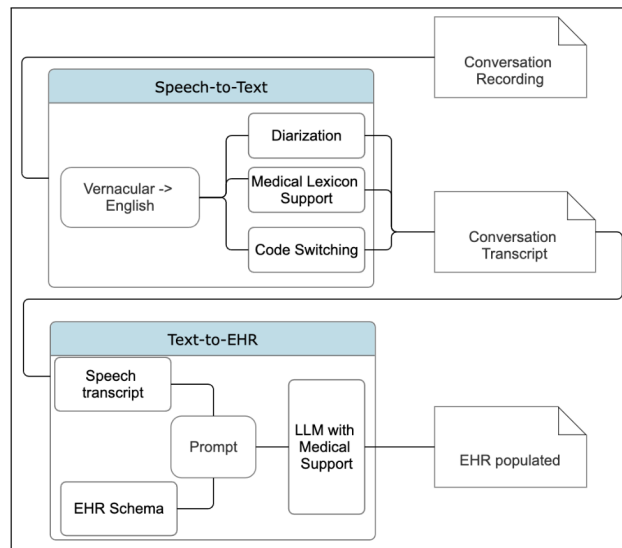


Figure 1: High and low-level Architecture Diagram

At the core of the system is the Speech-to-Text Conversion module, tasked with transforming conversations between doctors and patients, conducted in various regional languages, into English text. The converted English transcripts are then input into the EHR Population module, which

147 employs advanced large language models (LLMs) to systematically organize patient information
 148 within the EHR database. Additionally, we implement RAG to get corresponding ICD codes for
 149 diagnoses, NDC codes for medications, and CPT codes for procedures performed on the patient. The
 150 Text-to-EHR module includes sub-components focused on schema development and an LLM tailored
 151 for medical contexts, ensuring the EHR database is populated effectively and accurately. To further
 152 illustrate the flow of data as described above, we provide the following data-flow diagram:

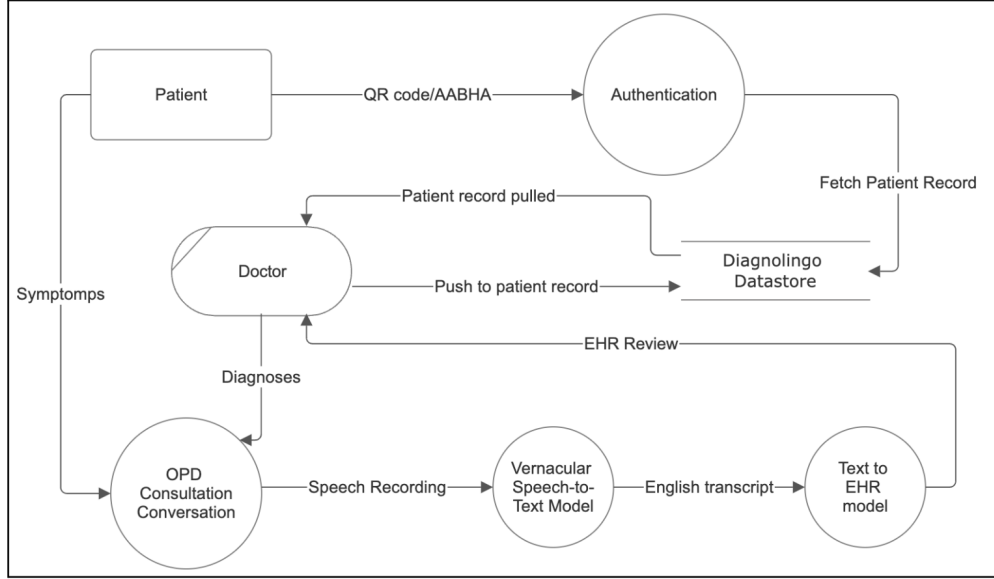


Figure 2: Data flow diagram

153 The speech data from its initial entry moves into the Speech-to-Text Conversion module, where
 154 it’s transformed into text transcripts. Subsequently, these transcripts are processed by the EHR
 155 Population module, which is responsible for structuring and entering patient information into the
 156 EHR database. An authentication mechanism is also incorporated to illustrate how patient records
 157 are securely retrieved from the Diagnolingo data repository, ensuring data access is based on verified
 158 patient information.

159 7 Experimental Design

160 7.1 Dataset

161 There are currently no datasets of medical doctor-patient conversations in any of the Indian vernacular
 162 languages/dialects. Given that we dont have the audio conversation itself we also obviously dont
 163 have the corresponding SOAP/EHR note for this conversation. Therefore, we leverage the ACI-
 164 bench [5] and MTS-Dialog [6] datasets as they contain the English transcript of a doctor-patient
 165 conversation and the corresponding EHR notes. For the audio data, I manually mocked and created
 166 the corresponding Hindi audio conversation using the English transcript available in ACI-bench/MTS-
 167 Dialog. This allows us to leverage the same SOAP/EHR notes in these databases to validate our
 168 generated notes.

- 169 1. **Medical vernacular audio conversation data:** None available, created via mocking. Given
 170 the time taken to create such data points, I started with mocking 3 conversations via a role-
 171 play with another person acting as a patient following the exact script given in ACI-bench
 172 database.
- 173 2. **Medical conversation transcript and EHR data:**

- 174 (a) **ACI-bench:** The ACI-Bench dataset is a benchmark designed specifically for Ambient
 175 Clinical Intelligence (ACI) systems, particularly those aimed at transforming doctor-
 176 patient conversations into structured clinical notes. It includes 1,007 dialogue instances
 177 with multiple physician-patient interaction types, focusing on realistic, complex ex-
 178 changes that are challenging for automatic summarization and clinical note generation.
 179 These dialogues are fully annotated, providing detailed mappings to clinical notes and
 180 making it suitable for use in developing and evaluating models that convert unstructured
 181 conversations into structured Electronic Health Record (EHR) notes. ACI-Bench was
 182 introduced as part of the 2023 MEDIQA challenge, which focused on dialogue-based
 183 summarization and clinical note generation
- 184 (b) **MTS-Dialog:** The MTS-Dialog dataset is a collection of 1,700 short doctor-patient
 185 conversations and corresponding summaries with normalized section headers including
 186 fam/sochx [FAMILY HISTORY/SOCIAL HISTORY], genhx [HISTORY of PRESENT
 187 ILLNESS], pastmedicalhx [PAST MEDICAL HISTORY], cc [CHIEF COMPLAINT],
 188 pastsurgical [PAST SURGICAL HISTORY], allergy, ros [REVIEW OF SYSTEMS],
 189 medications, assessment, exam, diagnosis, disposition, plan, edcourse [EMERGENCY
 190 DEPARTMENT COURSE], immunizations, imaging, gynhx [GYNECOLOGIC HIS-
 191 TORY], procedures, other history and labs.
- 192 3. **Non-medical vernacular data sources (Not currently useful):** Multilingual and code-
 193 switching ASR Challenge Dataset, NLTMP Pilot TTS Data For Indian Languages — Hindi,
 194 Punjabi, Tamil, And Indian English, Hindi ASR Challenge Data (ASR Speech Data Released
 195 Under 3rd Challenge) - NLTMP
- 196 4. **Medical Lexicons and Terminology Datasets:** Curated from medical dictionaries provided
 197 by ICD-10 (International Classification of Diseases, Tenth Revision), and SNOMED CT
 198 (Systematized Nomenclature of Medicine Clinical Terms). These were used to create the
 199 vector indexes for RAG.

200 7.2 Machine Learning Models/Algorithms/Pipeline

201 In developing Diagnolingo, we have chosen OpenAI’s Whisper [10] model to handle the speech-to-
 202 text conversion, given its impressive performance across languages, dialects, and accents. Whisper is
 203 particularly suitable for Indian outpatient settings due to its robustness with multilingual transcription,
 204 including Indian English and regional languages. This adaptability is crucial in India’s diverse
 205 linguistic landscape, where patients and providers often switch between languages mid-conversation
 206 or use a unique blend of vernacular terms.

207 For the Electronic Health Record (EHR) population module, we’ve implemented a Retrieval-
 208 Augmented Generation (RAG) [11] model, fine-tuned on medical data to accurately extract and
 209 structure critical information from these transcriptions. RAG combines retrieval and generation
 210 capabilities, allowing the model to dynamically pull from a pre-established repository of medical
 211 knowledge, including ICD codes for diagnoses, NDC codes for medications, and CPT codes for
 212 procedures, to ensure a high level of specificity and accuracy. This is particularly beneficial in an
 213 Indian outpatient context, where rapid conversational exchanges are the norm, and real-time accuracy
 214 is paramount.

215 Additionally, the custom EHR schema was designed specifically to align with common outpatient
 216 workflows in India, ensuring that it captures all relevant clinical details while being easy to use
 217 for healthcare providers. Together, these components—Whisper for transcription and a fine-tuned
 218 RAG model for EHR entry population—make Diagnolingo an effective solution for capturing and
 219 structuring healthcare data in India’s unique clinical landscape, supporting efficient and accurate
 220 medical documentation.

7.3 Evaluation Metrics

The evaluation of the speech-to-text module is based on the Word Error Rate (WER) and Character Error Rate (CER), which provide insights into the transcription accuracy. For the medical term recognition and information extraction tasks, precision, recall, and F1-score are utilized to measure the effectiveness of the NER component in accurately identifying and categorizing medical entities from the text.

In the context of the EHR population, the quality of data structuring is assessed through manual reviews and feedback from medical professionals, ensuring the system's outputs align with clinical expectations and needs. The qualitative feedback was taken from an MD for 5 samples generated on a scale of 1-5, along with detailed comments if any. The MD was tasked with looking at the conversation transcript and the final EHR note and then scoring how well the EHR note was able to capture the information in the transcript succinctly. A 5-point rating scale was used, where "1" represented "Poor," "2" indicated "Needs improvement," "3" was "Satisfactory," "4" signified "Good," and "5" denoted "Excellent." We also computed medical concepts recall, precision, and f-1 scores between the ACI-bench note and the generated EHR. In the future, when evaluating the system's overall performance, user satisfaction and the system's impact on reducing documentation time and improving diagnostic accuracy will also be considered, providing a holistic view of Diagnolingo's efficacy in a real-world healthcare setting.

8 Test Design

The test design for Diagnolingo is structured to evaluate the system's performance across its critical components: the Speech-to-Text Conversion module and the EHR Population module.

8.1 Procedure to Test the Speech-to-Text Conversion Module

1. Dataset Preparation: Utilize a subset of the dataset to serve as the test set. This subset should represent the linguistic diversity and medical complexity of the broader dataset.
2. Testing Execution: Run the Speech-to-Text Conversion module on the test set. The system will generate English transcripts from the multilingual audio recordings.
3. Results Comparison: Compare the generated transcripts against the transcripts in the test set. This comparison will focus on the accuracy of language transcription and the correctness of medical terminology.

8.2 Procedure to Test the EHR Population Module

1. Input Preparation: Provide the module with test transcripts that include a variety of medical discussions, ensuring they contain diverse entities and scenarios.
2. Population Execution: The EHR Population module processes the transcripts to identify and extract medical entities and relationships, populating the structured EHR according to the predefined schema.
3. Verification: Manually review a sample of the populated EHR entries against the transcripts to ensure the accuracy and completeness of the data capture and structuring. Get doctor reviews with domain specialization to confirm the findings. Finally, for a quantitative evaluation compare the EHR note generated by Diagnolingo with the EHR note provided in the ACI-bench/MTS-Dialog datasets.

8.3 Data Scheme of the Results

A detailed analysis will accompany the results, discussing the system's performance in various aspects, such as language handling, medical term accuracy, and EHR data structuring. Potential discrepancies or areas for improvement will be highlighted, providing a pathway for iterative enhancements to the system based on the test outcomes.

266 Through this detailed testing approach, Diagnolingo’s efficacy and readiness for deployment in a
267 real-world setting can be thoroughly assessed, ensuring that the system can reliably support healthcare
268 professionals in delivering efficient and accurate patient care.

269 **9 Deployment Model**

270 The deployment model for Diagnolingo is designed to ensure seamless integration into the existing
271 healthcare IT infrastructure, with a focus on scalability, security, and user accessibility. Our mobile-
272 application can be used in an outpatient doctor setting in India. The recorded audio is pushed to cloud
273 storage and the pipeline runs in the cloud. This allows the doctor to move forward with the next
274 conversation without having to wait until the previous results are compiled.

275 **9.1 Cloud-Based Infrastructure**

276 Diagnolingo will be deployed on a secure AWS cloud platform (which has inbuilt HIPAA compliance),
277 enabling scalable access to computational resources and data storage. This approach ensures that
278 the system can handle varying loads, from individual clinics to larger hospital networks, without
279 compromising performance.

280 **9.2 Modular Integration**

281 The system is designed with 3 modular components, speech-to-text, text-to-EHR, and predictive
282 analytics, allowing for flexible integration with existing Electronic Health Record (EHR) systems
283 and healthcare management tools. This modularity ensures that healthcare providers can choose to
284 adopt a certain component, eg, the speech-to-text only as well, easing the transition and encouraging
285 wider adoption. The predictive analytics is currently out of the scope of this capstone and will be
286 integrated at a later stage.

287 **9.3 Compliance and Security**

288 Deployment will adhere to stringent data protection and privacy regulations, ensuring that all patient
289 data is encrypted and securely processed. Compliance with national healthcare data standards and
290 regulations will be a priority, maintaining the trust of both healthcare providers and patients.

291 **9.4 User Training and Support**

292 To maximize the system’s impact, comprehensive training materials and support services will be
293 provided, ensuring that healthcare professionals can effectively utilize Diagnolingo. Continuous user
294 feedback will be encouraged to refine and enhance the system further. We might have to inculcate
295 certain practices amongst doctors, such as starting the recording only after all personal irrelevant
296 conversations are over with the patient to improve accuracy and support brevity.

297 **9.5 Phased Rollout**

298 Deployment will commence with a phased rollout, starting with pilot implementations in select
299 clinics and hospitals to gather real-world usage data and feedback. This approach allows for iterative
300 improvements before a broader rollout, ensuring that the system is robust and user-friendly at scale.

301 By adhering to this deployment model, we aim to become an integral part of India’s healthcare
302 ecosystem.

303 **10 Risks/Challenges**

304 The successful deployment and operation of Diagnolingo involve navigating a landscape of potential
305 risks and challenges across domain, technical, and business aspects.

10.1 Domain-Specific Risks

1. **Linguistic Diversity:** India's vast linguistic diversity poses a significant challenge, as the system must accurately transcribe and understand a multitude of regional languages and dialects, some of which may have limited digital resources available for training models.
2. **Medical Data Variability:** The variability in medical terminologies, procedures, and patient interactions across different regions and practices can affect the system's ability to consistently interpret and structure data.
3. **Regulatory Compliance:** Adhering to India's healthcare regulations and data privacy standards is crucial. Changes in regulations or non-compliance can pose legal risks and affect user trust.

10.2 Technical Risks

1. **Accuracy and Reliability:** Ensuring high accuracy in speech-to-text conversion and data extraction in a medical context is challenging. Any inaccuracies can lead to significant issues in patient care and record-keeping.
2. **Integration Complexity:** Seamlessly integrating Diagnolingo with existing EHR systems and healthcare workflows without disrupting current operations is complex and requires careful planning and execution.
3. **Scalability:** As the system scales, maintaining performance, managing data growth, and ensuring consistent service quality across a larger user base and data volume will be challenging.

10.3 Business Risks

1. **User Adoption:** Gaining widespread acceptance and usage by healthcare professionals, who may be accustomed to traditional methods of record-keeping, could be challenging. Resistance to change and varying levels of tech-savviness among users can impact adoption rates.
2. **Economic Viability:** The project must demonstrate clear value to justify the investment. Balancing cost-effectiveness with advanced technological features, especially in a cost-sensitive market like India, is crucial.
3. **Market Competition:** The healthcare technology sector is rapidly evolving. Keeping Diagnolingo competitive and innovative in the face of new entrants and advancements in AI and healthcare technologies is a significant business challenge.
4. **Data Security and Privacy:** Ensuring the utmost security and privacy of patient data is critical. Any breaches or lapses can severely damage reputation, user trust, and could have legal repercussions.

11 Tools and Dependencies

The Diagnolingo system leverages a suite of tools, libraries, and services. The system will be developed using Python (3.9) due to its widespread adoption in the data science and machine learning communities, and extensive library support and a robust ecosystem.

- **OpenAI Whisper (v2):** This is the core library for the speech-to-text conversion module. Whisper is selected for its state-of-the-art performance in diverse language and acoustic environments, crucial for handling India's linguistic diversity.
- **OpenAI GPT-4.0 mini:** GPT-4.0 mini model is a smaller and more cost and time efficient variant of OpenAI's GPT-4. We leverage this model for our transcript-EHR module and perform RAG over "ada" embeddings provided by OpenAI.

- **React (18.2.0):** Used to create reusable UI components. It optimizes rendering efficiency, allowing it to handle dynamic, real-time data changes smoothly, which is ideal for our web application.
- **FastAPI (0.101.0):** Used for backend API development, FastAPI is chosen for its simplicity and efficiency in creating web services, facilitating easy integration with front-end interfaces and other system components. FastAPI is built on ASGI (Asynchronous Server Gateway Interface), allowing it to handle asynchronous requests which are very useful in our use-case.
- **Cloud Computing Platform:** AWS was chosen on factors of cost and in-built HIPAA compliance. It will provide the computational resources and scalability for both the development and deployment phases.

12 Results

12.1 Speech-to-Text

For the mock Hindi audio to transcript module (using Whisper), we compute word error rate, character error rate, and unigram and bigram similarities between the 3 generated and original transcripts. Precision, recall, and F-1 scores of medical concepts between the generated transcript and the original transcript are also computed to ensure we are not missing important clinical information. Table 1 details all results for these 3 samples.

Table 1: Speech-to-Text module evaluation results (both medical and non-medical). WER is the Word Error Rate, and CER is the Character Error Rate. These values have been computed for the 1st, 3rd and 5th samples from the ACI-Bench database.

Index	Non-medical Metrics				Medical Metrics		
	WER	CER	Unigram Similarity	Bigram Similarity	Precision	Recall	F1
1	0.060	0.001	0.295	0.112	0.900	0.818	0.857
3	0.086	0.005	0.342	0.133	0.923	0.857	0.889
5	0.087	0.001	0.291	0.088	0.889	0.889	0.889

The average medical concept recall (0.854) and precision (0.903) values are very convincing even though there is language switching taking place in the module. Non-medical metrics such as average WER (0.07), CER (0.002), unigram similarity (0.31), and bigram similarity (0.11), are also convincingly good, and this robustness can build trust for both the patient and provider, which is paramount to our success.

12.2 Text-to-EHR

To convert the transcript to a SOAP note, we used the following prompt:

Write as a professional medical scribe, ensuring medical accuracy, clarity, and brevity. Go through the following doctor-patient conversation and create a SOAP note for it. A SOAP note consists of Subjective, Objective, Assessment and Plan sections. Just include these 4 sections and nothing else in the note. For each subfield in each of the four sections return a list of items in decreasing order of importance. If you do not have information for a particular field return an empty list. 1. 'Subjective' section includes items taken during the patient's verbal exam. Include 'Chief complaint', 'History of present illness', and 'Past social history' as subfields. 2. 'Objective' section includes findings from the physical examinations and diagnostics taken prior to the visit, including laboratory or imaging results, broken down by exam type. It should have the following subfields: 'Vital signs', 'Physical exam findings', 'Laboratory data', 'Imaging results', and 'Other diagnostic data'. If a specific exam type is not mentioned, return an empty list for that subfield. 3. 'Assessment' includes the doctor's diagnosis as a list in the subfield 'Diagnosis' in decreasing order of importance. 4. 'Plan'

386 section contains planned 'Tests', 'Referrals', 'Medications' along with 'Instructions' as separate
387 subfields. 'Medications' should contain a list of prescribed medications with a dictionary for each
388 containing medication 'Name', 'Dosage', 'Route', and 'Frequency'. If no medication is mentioned,
389 the 'Medications' should return an empty list. If you do not have data for a particular section or a
390 sub-section, return an empty list for that particular subfield. Ensure that the medical terminology used
391 in the conversation is accurately reflected. SOAP note should be concise, and avoid adding details
392 not explicitly mentioned in the conversation. Format the output as JSON with the keys: 'Subjective',
393 'Objective', 'Assessment' and 'Plan'. For each of these sections create dictionaries within for the
394 different subfields.

395 Conversation: {conversation}

396

397 Then RAG was performed over the vector databases for ICD, NDC, and CPT codes using the
398 following template:

399 Instruction: Only use the following returned documents to get the {ICD/NDC/CPT} codes. Ensure
400 matching is case-insensitive. If there are multiple codes possible, return all of them. If no useful
401 codes are found, just return 'N/A'. Output the codes as a comma-separated list in order of confidence
402 without spaces. Here are the relevant documents: {documents}

403

404 The results of executing this pipeline over 3 examples from the ACI-bench database are given in
405 Table 2. After analyzing the results, it is clear that for some tests the accurate CPT codes are hard to
406 fetch as their descriptions are not matching a lot. Therefore, for a lot of tests, the RAG doesn't return
407 any CPT codes resulting in missing field entries. However, on comparing the conversation with the
408 note generated, we can see a perfect recall of all medical concepts and terms used and that too in a
409 well-arranged dictionary format as intended.

410 In Table 3 we report the physician's feedback for 5 EHR samples we created using the ACI-bench
411 samples. The feedback is highly positive, with a mean rating of 3.8/5 for just the very first iteration
412 of the product. We now plan to improve our EHR note creation by assimilating the key points of the
413 feedback into our prompting strategy. This also serves as excellent error analysis as we get to look at
414 our outputs from the lens of the actual end user.

415 Newx, similar to the transcript generation module, we compute the precision, recall and F-1 scores
416 for medical concepts between the ACI-Bench SOAP notes and the generated EHRs. Considering that
417 our EHRs are by design much more succinct to improve faster readability, we expect a lower value
418 here. We were able to get an average precision of 0.526, recall of 0.533, and F-1 score of 0.527 across
419 the five pairs. This is slightly lower than expected, and in further iterations, with doctor feedback
420 incorporation, we expect these numbers to boost up as well.

Table 2: Output of EHR from Diagnolingo for 3 sample conversations from ACI-bench

ACI-conversation	Diagnolingo-EHR
<p>[doctor] hey george how are you today i understand you're here for some numbness and tingling in your fingers and some pain in your wrist</p> <p>[patient] right my left wrist and hand has been bothering me probably for a few months now with pain and numbness</p> <p>[doctor] okay and you said that's been ongoing for several months do you know what caused this type of pain or is it just something that started slowly or</p> <p>[patient] it just kinda started on it's own it i notice it mostly at night</p> <p>[doctor] okay</p> <p>[patient] sometimes it will i'll wake up and my hands asleep and i got ta shake it out</p> <p>[doctor] shake it out and okay[patient] and then some</p> <p>[doctor] what kind of work do you do</p> <p>[patient] i do yard work</p> <p>[doctor] yard work</p> <p>[patient] landscaping landscaping</p> <p>[doctor] landscaping okay so a lot of raking a lot of digging so a lot of repetitive type movements</p> <p>[patient] yeah it's pretty heavy labor but it's yeah the same thing day in and day out</p> <p>[doctor] okay okay just a couple questions for you you did say that you have the pain at night in that and you have to you get that numbness into the hand is it in all the fingers</p> <p>[patient] yeah it seems to happen to all my fingers but i notice it more in my thumb and pointer finger</p> <p>[doctor] okay okay and anything into that little into your fifth finger your little finger any numbness there at times no</p> <p>[patient] sometimes yeah it seems like it's numb too[doctor] okay what about your right hand any problems with that hand</p> <p>[patient] no i do n't seem to have any problems with my right hand so far it's just mostly my left</p> <p>[doctor] okay okay good and just a couple you know do you how do you have many or do you drink often do you have you know many any alcohol consumption</p> <p>[patient] i drink usually a a beer or two on fridays and saturdays on the weekends</p> <p>[doctor] okay and do you have any evidence of any anybody ever said that you had some rheumatoid arthritis in your hand or wrist anything like that</p> <p>[patient] no nobody say anything like that so i mean</p> <p>[doctor] okay okay good so let me go ahead and do a physical exam here real quick and you know i'm gon na quickly just listen to your heart and lungs okay that's good i'd like you to squeeze i'm gon na hold your hands here and i'd like you to squeeze both hands</p> <p>[patient] okay[doctor] you seem a little bit weaker on that left hand is that what you've noticed</p> <p>[patient] yeah i i i experienced some weakness in my left hand[doctor] okay do you you find that you're dropping things when you're picking it up is it to that level or</p> <p>[patient] yeah i drop things mostly because i have a hard time feeling it</p> <p>[doctor] okay okay good and so you you do have a a grip strength is less on the left and i just wan na touch your fingers here on the on the right side you can feel me touching all the fingers on the right</p> <p>[patient] yeah i can i can say you touch me but it feels a little more weird on the thumb side than my pointer finger side</p> <p>[doctor] okay okay and i wan na turn your wrist over here and turn your hand over and i'm gon na go ahead and tap on the right wrist on the back here does that do anything when i do that</p> <p>[patient] i still i feel a little jolt or a zing in my finger tips[doctor] okay and then when i do that on the left side</p> <p>[patient] yeah same thing</p> <p>[doctor] same thing okay so you do have a bilateral positive tinel's sign so so here's here's where i'm at i think your your diagnosis is beginning to have some bilateral carpal tunnel syndrome usually</p> <p>we see that with repetitive actions such as the landscaping the heavy labor and you you know your your clinical exam and and history sound like it's a carpal tunnel syndrome i do want to order so where are we gon na go from here i would like to order a a study it's called an emg where it it measures some of that electrical impulses down into your fingers we will follow up with that but as far as your treatment so the treatment for carpal tunnel syndrome is really some activity modification now i know you are a landscaper is there any way that you could be work to have some lighter work during the time</p> <p>[patient] i suppose i could try to pass it off to some of my other employees and delegate</p> <p>[doctor] okay that would be good so that's i i just want you to kinda eliminate that the active repetitive motions that you're doing all the time just for a couple weeks i'm also gon na give you a wrist splint to wear and that should help and i'd like you to take ibuprofen six hundred milligrams every six hours and then i wan na see you back here in the office in two weeks and in that two week period i think we're gon na see if there's need for any other intervention</p> <p>[patient] yeah it sounds like a good first start</p> <p>[doctor] okay take care and i'll look forward to see you in two weeks</p> <p>[patient] very good appreciate your time</p>	<p>{ 'Subjective': { 'Chief complaint': ['Numbness and tingling in left fingers', 'Pain in left wrist'], 'History of present illness': ['Symptoms have been ongoing for a few months', 'Noticed mostly at night', 'Wakes up with hands asleep, needs to shake them out', 'Weakness in left hand, difficulty feeling objects, drops things'], 'Past social history': ['Works in landscaping, involving heavy labor and repetitive movements', 'Occasional alcohol consumption (1-2 beers on weekends)'] }, 'Objective': { 'Vital signs': [], 'Physical exam findings': ['Weaker grip strength in left hand', 'Bilateral positive Tinel's sign', 'Altered sensation in left thumb and pointer finger'], 'Laboratory data': [], 'Imaging results': [], 'Other diagnostic data': [] }, 'Assessment': { 'Diagnosis': ['Bilateral carpal tunnel syndrome'] }, 'Plan': { 'Tests': ['Order EMG to measure electrical impulses in fingers'], 'Referrals': [], 'Medications': [{ 'Name': 'Ibuprofen', 'Dosage': '600 mg', 'Route': 'Oral', 'Frequency': 'Every 6 hours' }], 'Instructions': ['Activity modification: reduce repetitive motions', 'Wear wrist splint', 'Follow up in 2 weeks to reassess and consider further intervention'] }, 'Codes': { 'Tests': { 'Order EMG to measure electrical impulses in fingers': [] }, 'Medications': { 'Ibuprofen, 600 mg': ['182181089', '904585361', '904585461'] }, 'Diagnosis': { 'Bilateral carpal tunnel syndrome': ['G5603', 'G5600'] } } }</p>

Table 2 continued from previous page

ACI-conversation	Diagnolingo-EHR
<p>[doctor] hi billy how are you what's been going on the medical assistant told me that you're having some difficulty urinating</p> <p>[patient] yeah yeah i i did n't really wan na come in to talk about it's kinda weird but i think probably over the last six months i'm just not peeing right it just does n't seem to be normal</p> <p>[doctor] okay and do you feel like you're emptying your bladder fully or do you feel like you still have some urine left in there when you when you finish</p> <p>[patient] most of the times i'm okay but sometimes if i stand there long enough i i can kinda go a little bit more so it's taking a while actually to just go to the bathroom</p> <p>[doctor] okay and are you waking up at night to go to the bathroom does it impact your sleep</p> <p>[patient] yeah i try to empty my bladder now right before i go to bed and and not drink anything but i'm still probably getting up three or four times a night to go to the bed</p> <p>[doctor] okay so you're getting up about three or four times a night and and how long has this been going on you said for about six months</p> <p>[patient] yeah six months to like this and it's probably been a little bit worse over the last six months and maybe it's been longer i just did n't want to bring it up</p> <p>[doctor] okay so you think it's been going on longer okay alright now how about have you had any burning when you urinate at all</p> <p>[patient] no it i do n't think it burns[doctor] no burning when you urinate okay and and any other any other issues any problems with your bowels any constipation issues</p> <p>[patient] hmmm no i i had diarrhea last week but i think i ate something bad</p> <p>[doctor] okay and ever have you ever had any issues where you had what we call urinary retention where you could n't pee and you needed to have like a catheter</p> <p>[doctor] alright and have you ever seen a urologist i do n't think so you've been my patient for a while i do n't remember ever sending you but have you ever seen one</p> <p>[doctor] okay now tell me how are you doing with your with your heart when was the last time you saw doctor moore the cardiologist i know that you had the the stent placed in your right coronary artery about what was that twenty eighteen</p> <p>[patient] yeah sounds about right i think i just saw him in november</p> <p>[doctor] he said everything was okay alright and so you have n't had any chest pain or shortness of breath you're still walking around doing your activities of daily living are you exercising</p> <p>[doctor] let's talk a little bit about your diabetes how are how are you doing with your sugars are you watching your diet. okay and are you still taking the metformin [patient] yep</p> <p>[doctor] you are okay alright now i wan na go ahead and just move on to a quick physical exam okay i'm gon na be calling out some of my exam findings and i'm gon na let you know what that means</p> <p>when i'm done okay alright i do have to do a rectal exam i apologize i'm just gon na be calling it out what what i what i appreciate okay so on your heart exam i do appreciate a slight three out of six</p> <p>systolic ejection murmur hurt at the left base on your lung exam your lungs are clear to auscultation bilaterally on your abdominal exam your abdomen is nontender and nondistended i do n't appreciate</p> <p>any masses or any rebound or guarding on your prostate exam i do appreciate an enlarged prostate i do n't appreciate any masses on physical exam so what what does that mean billy so that ultimately</p> <p>means that you know everything looks good you know you have that little heart murmur which i believe you you've had in the past but we're gon na go ahead and look into that you know your prostate</p> <p>seems a little bit enlarged to me on physical exam so let's talk about how we can go about and and remedy that okay so for your first problem of this you know difficulty urinating i wan na go ahead</p> <p>and just order some routine labs i wan na get a a psa that kind of that ultimately kinda looks for prostate cancer issues which i do n't think you have because we did n't really appreciate that on physical</p> <p>exam i wan na go ahead and we can try to start you on what we call flomax zero point four milligrams once a day you should take it at night because it can cause people to get a little bit dizzy if they</p> <p>take it in the morning so i would take it at night and i wan na go ahead and refer you to a urologist just to look into this more so we can go ahead and and get this problem solved for you okay i'm also</p> <p>gon na go ahead and just order some routine blood tests just to make sure that we are not missing anything do you have any questions about that and i wan na go ahead and order a urinalysis and a urine culture</p> <p>[doctor] flomax and that's why we are gon na refer you to the urologist so that they can go ahead and talk to you about you know the most current treatment options for you okay. alright for your second problem of your coronary artery disease i wan na go ahead and order an echocardiogram just to follow up on that heart murmur that you had and i wan na go ahead and</p> <p>continue you on the lipitor forty milligrams a day and the aspirin and the metoprolol and i wan na go ahead and order a lipid panel any questions about that</p> <p>[doctor] okay and then for your third problem of your diabetes it sounds like you're doing really well let's go ahead and continue you on the metformin a thousand milligrams twice a day we will</p> <p>go ahead and order a hemoglobin a1c to see if we need to make any adjustments to that and i'm gon na see you again in about three to four weeks okay i want you to call me or message me in the patient portal if you have any concerns</p> <p>[patient] alright when is the urologist gon na call me</p> <p>[doctor] i'm gon na reach out i'm gon na reach out to them now and see if they can get you in this week</p> <p>[patient] sounds good</p> <p>[doctor] okay alright well great it was good to see you bye</p>	<p>{ 'Subjective': { 'Chief complaint': ['Difficulty urinating for the past six months'], 'History of present illness': ['Weak urination stream', 'Feeling of incomplete bladder emptying', 'Nocturia (3-4 times per night)', 'No burning during urination', 'Recent diarrhea attributed to food'], 'Past social history': []], 'Objective': { 'Vital signs': [], 'Physical exam findings': ['Slight 3/6 systolic ejection murmur at left base', 'Lungs clear to auscultation bilaterally', 'Abdomen nontender and nondistended', 'Enlarged prostate without masses'], 'Laboratory data': [], 'Imaging results': [], 'Other diagnostic data': []], 'Assessment': { 'Diagnosis': ['Benign prostatic hyperplasia (BPH)', 'Coronary artery disease', 'Diabetes mellitus']}, 'Plan': { 'Tests': ['PSA test', 'Routine blood tests', 'Urinalysis', 'Urine culture', 'Echocardiogram', 'Lipid panel', 'Hemoglobin A1c'], 'Referrals': ['Urologist'], 'Medications': [{ 'Name': 'Flomax', 'Dosage': '0.4 mg', 'Route': 'Oral', 'Frequency': 'Once daily at night'], { 'Name': 'Lipitor', 'Dosage': '40 mg', 'Route': 'Oral', 'Frequency': 'Once daily'], { 'Name': 'Aspirin', 'Dosage': '', 'Route': 'Oral', 'Frequency': 'Once daily'], { 'Name': 'Metoprolol', 'Dosage': '', 'Route': 'Oral', 'Frequency': 'Once daily'], { 'Name': 'Metformin', 'Dosage': '1000 mg', 'Route': 'Oral', 'Frequency': 'Twice daily']}, 'Instructions': ['Follow up in 3-4 weeks', 'Contact via patient portal for concerns']}, 'Codes': { 'Tests': { 'PSA test': ['G0103'], 'Routine blood tests': [], 'Urinalysis': [], 'Urine culture': [], 'Echocardiogram': ['93320', '93321', '93304'], 'Lipid panel': ['80061', '83718', '84478', '82465'], 'Hemoglobin A1c': []], 'Medications': { 'Flomax, 0.4 mg': ['597005801', '904640161', '51079029420'], 'Lipitor, 40 mg': ['51079045620', '68084058901', '54868422900'], 'Aspirin, .': ['63739002401', '904404073', '63739052201', '17714001110'], 'Metoprolol,': ['186109039', '51079025520', '68084030301', '378003201'], 'Metformin, 1000 mg': ['50268053115', '51079097220']}, 'Diagnosis': { 'Benign prostatic hyperplasia (BPH)': ['N401', 'N400'], 'Coronary artery disease': ['I2584', 'I25700', 'I2081'], 'Diabetes mellitus': ['E0840', 'E1040', 'E088', 'E0841']}]}</p>

Table 2 continued from previous page

ACI-conversation	Diagnolingo-EHR
<p>[doctor] so anna good to see you today so reading here in your appointment notes you were you were diagnosed with kidney stones from your your pcp and you currently have one and so they they had you come in so can you tell me what happened how's all that going for you</p> <p>[patient] sure i've been having some back pain on my right side it's been lasting for about a week now and i also started to see some blood in my urine</p> <p>[doctor] okay so on the right side so does that pain does it move anywhere or is it just kinda stay in that that one area</p> <p>[patient] yeah it's moved down a little bit on to my right lower side a little bit</p> <p>[doctor] side okay so how would you describe the pain is it constant or is does it come and go</p> <p>[patient] it's pretty constant</p> <p>[doctor] okay did you notice any pain when you're urinating i know i know you say you you saw you see blood but any pain with that</p> <p>[patient] no no real pain when i'm when i'm peeing at all</p> <p>[doctor] okay so have you taken anything i know have you tried like azo or any of that to</p> <p>[patient] i took some ibuprofen that helped a little bit</p> <p>[doctor] okay</p> <p>[patient] but it still hurts even with ibuprofen</p> <p>[doctor] alright have you noticed any nausea vomiting fever chills</p> <p>[patient] i have n't thrown up but i felt a little bit nauseated</p> <p>[doctor] little nauseated yeah that's we expected so have you do you have a family history of kidney stones i know some people when they have them like their parents have them stuff but</p> <p>[patient] yeah my my dad had kidney stones i think he has passed a couple of them i'm not quite sure</p> <p>[doctor] alright and have you had any in the past or is this your first one</p> <p>[patient] this is my first time i've never had this before</p> <p>[doctor] okay alright so we'll do we'll do an exam on you just to check you out so i guess you were in pain and stuff over the over the easter easter break there that</p> <p>[patient] yeah yeah i had some pain over the weekend i saw my pediatrician this morning so they sent me over here they were concerned that i might have a kidney stone</p> <p>[doctor] a little bit alright so on your exam of your abdomen there is mild pain and tenderness to palpation of the abdomen there's no rebound or guarding there is cva located near your flank tenderness on the right so that means that everything looks good but you do have what seems to be some inflammation of your kidney okay so we were able to get a ct of your your side and it showed that you do have a stone measuring point five centimeters in size and it's located in the proximal right ureter and so that's that duck that passes from your kidney to your bladder alright i do n't see any evidence of hydronephrosis so that means that there's not obstructing ureter causing swelling in your kidney which is which is pretty good so let's talk a little bit about my assessment and plan so you do have that kidney stone on the right so what i'm recommending is i want you to push fluids just to help facilitate you passing that stone alright have you been taking in have you been drinking enough water do you think so far</p> <p>[patient] probably not enough i drink some but</p> <p>[doctor] okay yeah i i want you to to drink try drink as much as possible just to see if we can get you hydrated and pass the stone what i'm gon na do is i'm gon na prescribe you ibuprofen eight hundred milligrams you can take that as needed for pain i know you said you were in that much pain just in case it does start to move you're in pain i to take the ibuprofen i'm also gon na give you a strainer for you to strain your pee so we can see you wan na know when that that stone does pass gon na order a bmp and a urinalysis and a urine culture and based on what the urinalysis shows we can decide if i can decide if i need to put you on antibiotics if you do have an infection of some kind and i wan na see you back in about a week to two weeks and if you're still having symptoms we can discuss further treatment such as a lithotripsy and it's it's a mainly minimally invasive procedure where we use shock waves to try to break up that stone but otherwise do you have any other questions for me</p> <p>[patient] no i do n't think so</p> <p>[doctor] alright so we will see you back in a week or two and i'll have my nurse come in with that prescription and hopefully with all the treatment you'll be able to pass the stone okay alright</p>	<p>{'Subjective': {'Chief complaint': ['Back pain on the right side', 'Blood in urine'], 'History of present illness': ['Back pain lasting for about a week', 'Pain has moved to the right lower side', 'Constant pain', 'No pain during urination', 'Nausea but no vomiting', 'Family history of kidney stones (father)'], 'Past social history': ['First occurrence of kidney stones']},</p> <p>'Objective': {'Vital signs': ['No fever', 'Blood pressure and heart rate normal'], 'Physical exam findings': ['Mild pain and tenderness to palpation of the abdomen', 'CVA tenderness on the right', 'No rebound or guarding'], 'Laboratory data': [], 'Imaging results': ['CT scan shows a stone measuring 0.5 cm in the proximal right ureter', 'No evidence of hydronephrosis'], 'Other diagnostic data': []},</p> <p>'Assessment': {'Diagnosis': ['Kidney stone in the proximal right ureter', 'Mild kidney inflammation']},</p> <p>'Plan': {'Tests': ['BMP', 'Urinalysis', 'Urine culture'], 'Referrals': [], 'Medications': [{'Name': 'Ibuprofen', 'Dosage': '800 mg', 'Route': 'Oral', 'Frequency': 'As needed for pain'}],</p> <p>'Instructions': ['Increase fluid intake to facilitate passing the stone', 'Use a strainer to collect urine to monitor for stone passage', 'Follow up in 1-2 weeks for reassessment']},</p> <p>'Codes': {</p> <p>'Tests': {'BMP': [], 'Urinalysis': [], 'Urine culture': []},</p> <p>'Medications': {'Ibuprofen, 800 mg': ['182181089', '904585361', '904585461']},</p> <p>'Diagnosis': {'Kidney stone in the proximal right ureter': ['N202'], 'Mild kidney inflammation': ['N/A']}}}</p>

Table 3: Qualitative Doctor Feedback with Scores and Comments for 5 samples of EHR notes created

Diagnolingo-EHR	Physician Score	Physician Comments
<p>{'Subjective': {'Chief complaint': ['Numbness and tingling in left fingers', 'Pain in left wrist'], 'History of present illness': ['Symptoms have been ongoing for a few months', 'Noticed mostly at night', 'Wakes up with hands asleep, needs to shake them out', 'Weakness in left hand, difficulty feeling objects, drops things'], 'Past social history': ['Works in landscaping, involving heavy labor and repetitive movements', 'Occasional alcohol consumption (1-2 beers on weekends')]], 'Objective': {'Vital signs': []}, 'Physical exam findings': ['Weaker grip strength in left hand', 'Bilateral positive Tinel's sign', 'Altered sensation in left thumb and pointer finger'], 'Laboratory data': [], 'Imaging results': [], 'Other diagnostic data': []}, 'Assessment': {'Diagnosis': ['Bilateral carpal tunnel syndrome']}, 'Plan': {'Tests': ['Order EMG to measure electrical impulses in fingers'], 'Referrals': [], 'Medications': [{'Name': 'Ibuprofen', 'Dosage': '600 mg', 'Route': 'Oral', 'Frequency': 'Every 6 hours'}], 'Instructions': ['Activity modification: reduce repetitive motions', 'Wear wrist splint', 'Follow up in 2 weeks to reassess and consider further intervention']}, 'Codes': {'Tests': ['Order EMG to measure electrical impulses in fingers': []], 'Medications': ['Ibuprofen, 600 mg': ['182181089', '904585361', '904585461']], 'Diagnosis': ['Bilateral carpal tunnel syndrome': ['G5603', 'G5600']]}}</p>	3	<p>Chief complaint: should include the specific words "numbness & tingling," as that pins-and-needles feeling is the most clear sign indicating carpal tunnel. Also the distribution of the pain on his hand is important (thumb + index finger). He mentions that he feels the numbness and tingling mostly in his thumb and index finger. This is the part of the hand that is most affected in carpal tunnel, so this info helps determine whether he has carpal tunnel or a similar but different condition (e.g., cubital tunnel, which would affect his little finger more). Also, should ideally include the fact that he denies any past history of rheumatoid arthritis. Doctors ask questions like this to rule out other possible conditions and be more sure.</p>
<p>{'Subjective': {'Chief complaint': ['Difficulty urinating for the past six months'], 'History of present illness': ['Weak urination stream', 'Feeling of incomplete bladder emptying', 'Nocturia (3-4 times per night)', 'No burning during urination', 'Recent diarrhea attributed to food'], 'Past social history': []}, 'Objective': {'Vital signs': []}, 'Physical exam findings': ['Slight 3/6 systolic ejection murmur at left base', 'Lungs clear to auscultation bilaterally', 'Abdomen nontender and nondistended', 'Enlarged prostate without masses'], 'Laboratory data': [], 'Imaging results': [], 'Other diagnostic data': []}, 'Assessment': {'Diagnosis': ['Benign prostatic hyperplasia (BPH)', 'Coronary artery disease', 'Diabetes mellitus']}, 'Plan': {'Tests': ['PSA test', 'Routine blood tests', 'Urinalysis', 'Urine culture', 'Echocardiogram', 'Lipid panel', 'Hemoglobin A1c'], 'Referrals': ['Urologist'], 'Medications': [{'Name': 'Flomax', 'Dosage': '0.4 mg', 'Route': 'Oral', 'Frequency': 'Once daily at night'}, {'Name': 'Lipitor', 'Dosage': '40 mg', 'Route': 'Oral', 'Frequency': 'Once daily'}, {'Name': 'Aspirin', 'Dosage': '', 'Route': 'Oral', 'Frequency': 'Once daily'}, {'Name': 'Metoprolol', 'Dosage': '', 'Route': 'Oral', 'Frequency': 'Once daily'}, {'Name': 'Metformin', 'Dosage': '1000 mg', 'Route': 'Oral', 'Frequency': 'Twice daily'}], 'Instructions': ['Follow up in 3-4 weeks', 'Contact via patient portal for concerns']}, 'Codes': {'Tests': ['PSA test': ['G0103'], 'Routine blood tests': [], 'Urinalysis': [], 'Urine culture': [], 'Echocardiogram': ['93320', '93321', '93304'], 'Lipid panel': ['80061', '83718', '84478', '82465'], 'Hemoglobin A1c': []], 'Medications': {'Flomax, 0.4 mg': ['597005801', '904640161', '51079029420'], 'Lipitor, 40 mg': ['51079045620', '68084058901', '54868422900'], 'Aspirin,': ['63739002401', '904404073', '63739052201', '17714001110'], 'Metoprolol,': ['186109039', '51079025520', '68084030301', '378003201'], 'Metformin, 1000 mg': ['50268053115', '51079097220']}, 'Diagnosis': ['Benign prostatic hyperplasia (BPH)': ['N401', 'N400'], 'Coronary artery disease': ['I2584', 'I25700', 'I2081'], 'Diabetes mellitus': ['E0840', 'E1040', 'E088', 'E0841']]}}</p>	4	<p>Seems like the HPI part is missing here, which is the part that tells the "story" of why he is coming to the doctor. Should include both pertinent positives AND negatives. So it should include difficulty urinating for 6 months (perhaps longer), waking up 2-3 times at night to urinate, and denying any burning sensation associated with it. Irrelevant to his condition, but the HPI should also mention his diarrhea from last week that was probably caused by food. On physical exam, should note that abdomen is nontender and nondistended without any "rebound", so nothing going on in his abdomen. Also, most importantly, should mention the enlarged prostate on rectal exam</p>

Table 3 continued from previous page

Diagnolingo-EHR	Physician Score	Physician Comments
<p>{'Subjective': {'Chief complaint': ['Back pain on the right side', 'Blood in urine'], 'History of present illness': ['Back pain lasting for about a week', 'Pain has moved to the right lower side', 'Constant pain', 'No pain during urination', 'Nausea but no vomiting', 'Family history of kidney stones (father)'], 'Past social history': ['First occurrence of kidney stones']], 'Objective': {'Vital signs': ['No fever', 'Blood pressure and heart rate normal'], 'Physical exam findings': ['Mild pain and tenderness to palpation of the abdomen', 'CVA tenderness on the right', 'No rebound or guarding'], 'Laboratory data': [], 'Imaging results': ['CT scan shows a stone measuring 0.5 cm in the proximal right ureter', 'No evidence of hydronephrosis'], 'Other diagnostic data': []}, 'Assessment': {'Diagnosis': ['Kidney stone in the proximal right ureter', 'Mild kidney inflammation']}, 'Plan': {'Tests': ['BMP', 'Urinalysis', 'Urine culture'], 'Referrals': [], 'Medications': [{'Name': 'Ibuprofen', 'Dosage': '800 mg', 'Route': 'Oral', 'Frequency': 'As needed for pain'}], 'Instructions': ['Increase fluid intake to facilitate passing the stone', 'Use a strainer to collect urine to monitor for stone passage', 'Follow up in 1-2 weeks for reassessment']}, 'Codes': {'Tests': {'BMP': [], 'Urinalysis': [], 'Urine culture': []}, 'Medications': {'Ibuprofen, 800 mg': ['182181089', '904585361'], '904585461'}}}, 'Diagnosis': {'Kidney stone in the proximal right ureter': ['N202'], 'Mild kidney inflammation': ['N/A']}}}</p>	3	<p>Mention that patient tried ibuprofen which helped a little bit, and also should mention no complaints of fever or chills (i.e., no fever, so it's not an infection). On physical exam, it might be good to mention "flank pain", but this is essentially the same as the CVA pain ("costovertebral angle" pain).</p>
<p>{'Subjective': {'Chief complaint': ['Evaluation of kidney stones'], 'History of present illness': ['Sudden onset of right back pain about a week ago while working with horses', 'Pain radiated to the groin but not completely', 'Nausea and sweating during the episode', 'No current pain, but had intermittent pain after diagnosis', 'Straining urine with no visible stones or blood'], 'Past social history': ['No previous history of kidney stones', 'Family history of kidney stones (father)'], 'Objective': {'Vital signs': ['Blood pressure and heart rate normal', 'No fever'], 'Physical exam findings': ['Tenderness on the right side (CVA tenderness)', 'No abdominal tenderness, rebound, or guarding'], 'Laboratory data': ['CT scan shows a 0.5 cm stone in the proximal right ureter', 'No evidence of hydronephrosis'], 'Imaging results': [], 'Other diagnostic data': []}, 'Assessment': {'Diagnosis': ['Kidney stone in proximal right ureter', 'No hydronephrosis']}, 'Plan': {'Tests': ['BMP', 'Urinalysis', 'Urine culture'], 'Referrals': [], 'Medications': [{'Name': 'Oxycodone', 'Dosage': '5 mg', 'Route': 'Oral', 'Frequency': 'Every 6 to 8 hours'}], 'Instructions': ['Push fluids to facilitate stone passage', 'Continue using urine strainer', 'Follow up in 1-2 weeks to reassess']}, 'Codes': {'Tests': {'BMP': [], 'Urinalysis': [], 'Urine culture': []}, 'Medications': {'Oxycodone, 5 mg': ['54368663', '66591091332']}, 'Diagnosis': {'Kidney stone in proximal right ureter': ['N202'], 'No hydronephrosis': ['N/A']}}}</p>	5	<p>This one looks good, covered all of the important stuff pretty consistently</p>
<p>{'Subjective': {'Chief complaint': ['Right knee pain after jumping on a trampoline'], 'History of present illness': ['Injury occurred four days ago with a popping sensation and subsequent swelling', 'Pain is significant when standing, but weight-bearing is possible', 'Tried ibuprofen and ice with minimal relief'], 'Past social history': ['Diet is generally good, but patient forgets to check blood sugar levels regularly']}, 'Objective': {'Vital signs': [], 'Physical exam findings': ['Swelling noted in the right knee', 'Pain on palpation of the inside of the knee', 'Limited range of motion with pain on flexion and extension', 'Negative McMurray's test'], 'Laboratory data': [], 'Imaging results': [], 'Other diagnostic data': []}, 'Assessment': {'Diagnosis': ['Injury to the medial collateral ligament (MCL) of the right knee', 'Diabetes management requiring follow-up']}, 'Plan': {'Tests': ['MRI of the right knee to assess for additional damage'], 'Referrals': [], 'Medications': [{'Name': 'Mobic', 'Dosage': 'Not specified', 'Route': 'Oral', 'Frequency': 'As needed for pain'}, {'Name': 'Metformin', 'Dosage': '500 mg', 'Route': 'Oral', 'Frequency': 'Continue as prescribed'}], 'Instructions': ['Use a straight leg brace for the right knee', 'Ice the knee once an hour for 15 minutes', 'Recheck hemoglobin A1c']}, 'Codes': {'Tests': ['MRI of the right knee to assess for additional damage': []], 'Medications': {'Mobic, Not specified': ['597002901'], 'Metformin, 500 mg': ['50268053115', '51079097220']}, 'Diagnosis': ['Injury to the medial collateral ligament (MCL) of the right knee': ['S83411D', 'S83411S', 'M23631', 'S83419D'], 'Diabetes management requiring follow-up': ['N/A']}}}</p>	4	<p>Only thing I'd add is the specific description of where it hurts inside her knee ("kind of that inside part of my kneecap"). It also might be relevant to mention the past medical history in this patient (diabetes), since even though she came in for her knee they ended up taking care of her diabetes and diet as well.</p>

13 Error Analysis

In our error analysis of the Text-to-SOAP-to-EHR pipeline, we evaluated how Retrieval-Augmented Generation (RAG) flagged and selected codes for ICD (diagnoses), NDC (medications), and CPT (procedures). Results for the first sample in the ACI-bench database are presented in Table 4. This analysis provided key insights into the nuances of each coding system and highlighted areas where RAG excels and where it could be further refined.

13.1 Diagnoses: ICD Code Selection and Hierarchical Nuances

In the case of diagnoses, the term "Bilateral carpal tunnel syndrome" appeared in the SOAP note. RAG retrieved four relevant ICD codes: G5603, G5600, G5601, and G5602. Of these, the first two codes, G5603 and G5600, referred to "Carpal tunnel syndrome in the upper limb" but did not specify which limb. However, G5603 was prioritized as the closest match, as it specifically included the term "bilateral," while G5600 was a general code for carpal tunnel syndrome in the upper limb without further specification. The other two codes, G5601 and G5602, represented carpal tunnel syndrome in the right and left limbs, respectively, and were thus less appropriate given the explicit mention of "bilateral" in the SOAP note. The final result prioritized G5603, followed by G5600, indicating the system's confidence in the match hierarchy.

To verify RAG's accuracy, we cross-checked the result in the ICD-10 directory, which confirmed that G5603 is indeed the ICD-10-CM code for carpal tunnel syndrome in both upper limbs. This consistency demonstrates that RAG performed well in identifying the correct codes and ordering them by relevance. This approach also offers flexibility for the clinician, who can either select the specific code (G5603) or opt for the more general G56, depending on confidence in the specificity needed. ICD-10 codes are inherently hierarchical, allowing for both specific and broader codes to accommodate different levels of diagnostic certainty. This example underscores the robustness of RAG in accurately capturing medical terminology context and suggests it can provide useful decision support in clinical scenarios.

13.2 Medications: Challenges with NDC Code Variety

When examining medications, a unique challenge emerged due to the structure of NDC (National Drug Code) assignments. The medication "Ibuprofen, 600 mg" retrieved four NDC codes for Ibuprofen, but each code corresponded to a different manufacturer. NDC codes are assigned to every drug formulation produced by a specific pharmaceutical company, meaning that even identical drugs like Ibuprofen from different manufacturers have unique NDC codes. This presents an immediate challenge: although the drug remains the same, the coding system does not accommodate a single code for all generic versions of Ibuprofen.

The variation in NDC codes highlights a real-world issue in medical coding. In practice, a doctor may not be concerned with which specific formulation of Ibuprofen was prescribed, as long as the active ingredient and dosage match. In this example, while each of the top four RAG results pointed to Ibuprofen, the minor differences in names, such as "Ibupro" or "Ibup," reflected only brand variations, not any difference in the actual drug. A quick database search confirmed the extent of this issue, revealing over 5,700 NDC codes for Ibuprofen alone. This multiplicity of codes creates a technical barrier to deployment in clinical settings, where the emphasis is more on the active ingredient than on the manufacturer.

To address this, future work could focus on a system to consolidate NDC codes based on the drug's chemical formulation, rather than company-specific codes. By indexing drugs according to their active ingredients and leveraging medical databases for comprehensive drug information, the pipeline could simplify medication coding. Such a strategy would provide more clinically relevant information for doctors and reduce the complexity involved in selecting the correct NDC code.

13.3 Procedures: CPT codes leveraging real-world information

The most compelling results from RAG came from analyzing procedure codes, particularly with test order, "Order EMG to measure electrical impulses in fingers." The RAG system initially returned two possible CPT codes, 9602 and 9603, both describing "Dynamic EMG tests." However, the final

471 output was “None,” which seemed puzzling at first glance. Further investigation into the nature of
472 EMG tests revealed that EMG for fingers is classified as a "Static" test, not a "Dynamic" one. As a
473 result, the first two codes, despite appearing relevant on the surface, were ultimately discarded by
474 RAG due to the nature of the test being ordered.

475 This outcome exemplifies why RAG is advantageous over simple keyword matching in medical
476 contexts. Rather than relying on a rigid matching process, RAG can integrate real-world information
477 and context, allowing it to identify nuances that may not be immediately apparent. This capability is
478 crucial in medical applications, where subtle differences in test types can have significant implications
479 for both diagnosis and billing. By filtering out codes based on specific requirements, such as static
480 versus dynamic testing, RAG demonstrates its potential as a valuable support tool that can enhance
481 coding accuracy in clinical environments. However, it also highlighted that we may want to consider
482 switching to a separate procedure coding system which is inclusive of all different types of procedures.

483 In conclusion, this error analysis of the Text-to-SOAP-to-EHR pipeline highlights both the strengths
484 and limitations of using RAG for medical coding across ICD, NDC, and CPT systems. By addressing
485 the limitations uncovered through this analysis, we can build a system that better aligns with clinical
486 workflows and meets the nuanced needs of real-world healthcare settings.

Table 4: Analyzing RAG output for medical codes for Diagnosis, Medications, and Procedures for Sample 1 from ACL-bench database.

Field	Name	RAG	Output
Diagnosis	Bilateral carpal tunnel syndrome	<p>### CODE: G5603 DESCRIPTION: Carpal tunnel syndrome; bilateral upper limbs</p> <p>### CODE: G5600 DESCRIPTION: Carpal tunnel syndrome; unspecified upper limb</p> <p>### CODE: G5602 DESCRIPTION: Carpal tunnel syndrome; left upper limb</p> <p>### CODE: G5601 DESCRIPTION: Carpal tunnel syndrome; right upper limb</p> <p>### DRUG: Ibuprofen NDC: 904585361</p>	G5603, G5600
Medication	Ibuprofen, 600 mg	<p>### DRUG: Ibuprofen NDC: 182181089</p> <p>### DRUG: ibuprofen NDC: 904585361</p> <p>### DRUG: Ibup NDC: 904585461</p> <p>### CODE: 96002 DESCRIPTION: Dynamic surface emg</p> <p>### CODE: 96003 DESCRIPTION: Dynamic fine wire emg</p> <p>### CODE: 97032 DESCRIPTION: Electrical stimulation</p> <p>### CODE: G0283 DESCRIPTION: Elec stim other than wound</p>	182181089, 904585361, 904585461
Procedure	Order EMG to measure electrical impulses in fingers		-

14 Discussion

In the developmental journey of Diagnolingo, the experimental process played a pivotal role in shaping the system's capabilities, particularly focusing on the Speech-to-Text conversion and EHR population modules. The breakdown of the model into two independent components and developing each of them using different language modeling techniques allows robustness as well. The four-step structured approach to evaluating these modules offers valuable insights into the system's potential effectiveness and areas for future enhancement. The first testing step was to compute the word error rate and character error rate between the generated and original transcripts. The second was to compute the precision and recall for medical concepts between the two transcripts. Then for the EHR generation, we tested qualitatively (third) using doctor feedback and quantitatively (fourth) using medical concepts precision and recall. These 4 steps allow a complete model evaluation and allow us to perform error analysis at each step (described in the previous section).

The results show the success of such an approach, as we were able to most importantly get a good average rating (3.8/5.0) from the physician looking at the generated notes. Our numbers for each of the 3 other quantitative evaluations also seem promising and have great potential to be refined further as we keep iterating based on feedback. The speech-to-text module especially has very good average recall (0.854), precision (0.903), and F-1 score (0.878) for medical terms across transcription. The average word (0.07) and character (0.002) error rates are also very low and build robustness into the system for downstream EHR formation.

The ongoing development and testing phase is integral to refining the system, ensuring it can effectively reduce the administrative load on healthcare providers and enhance the accuracy and accessibility of patient medical records. For a project like Diagnolingo with a lot of regulatory and adoption headwinds we highlight and employ a continuous, iterative development process, where feedback and testing insights will progressively shape Diagnolingo into a robust solution tailored to the unique challenges of the Indian healthcare landscape.

15 Lessons Learned and Reflections

Reflecting on the journey from the vision document in the Spring semester to the culmination of this report by the end of Fall semester has been an enlightening experience. The vision document set the foundation, outlining ambitious yet clear objectives and strategies to navigate the complexities of introducing a technologically advanced solution to small outpatient clinics in India—a sector traditionally reliant on paper-based records. This foundational phase was crucial in establishing a roadmap for what Diagnolingo aimed to achieve, considering the nuanced needs of its intended users and the specificities of the healthcare environment. Moving forward, the development of other reports required a deep dive into the technical and operational intricacies of the project. The transition from theory to application involved a rigorous examination of technical choices, evaluation frameworks, and deployment strategies, ensuring that each element of Diagnolingo was not only technically sound but also practically viable and aligned with the overarching goal of the project. Given that I was the only person on this project, I was shouldering the dual responsibilities of creator and critic, necessitating a continuous cycle of assessment and refinement. Looking back, one of the key lessons learned is the importance of time allocation for ideation and the potential benefits of more structured brainstorming phases. Even in solo projects, defining clear roles and responsibilities for different segments of the work can enhance efficiency and accountability, providing a clearer structure and timeline for the project's progression.

16 Future Work

After the fall term and the capstone project itself, I will focus on developing, refining and integrating its core components, aiming to solidify the foundation laid in this semester. I will focus on adding the predictive analytics modules, culminating in a rigorous testing phase to ensure seamless functionality and inter-module compatibility. This period is anticipated to present challenges, particularly in ensuring that these diverse components communicate effectively without glitches, potentially requiring additional resources in the form of sophisticated integration testing tools. Following this, I will shift towards iterative improvements based on doctor feedback and finalize application's features, setting the stage for developing a robust go-to-market strategy and exploring funding avenues by mid next

year. This systematic approach is designed to transition Diagnolingo from a promising prototype to a market-ready solution.

17 Conclusion

The journey of Diagnolingo has been a methodical and insightful endeavor aimed at transforming the landscape of healthcare documentation in India through an AI-powered EHR system. From conceptualizing a vision to addressing the nuanced demands of speech-to-text conversion and EHR population, the project has laid down a robust foundation for a system that promises to enhance the efficiency and accuracy of patient care in outpatient clinics. This upcoming phase is pivotal, as it will transition the project from a developmental stage to a pre-market phase, setting the stage for real-world application and further innovation in the healthcare technology domain. The most critical result for us is that "the system works," and the doctors across the board confirmed the value of getting such a product in their hands. While the exact metrics will continue improving with time as we iterate on model development, the immediate next step for us is to get the prototype in the hands of multiple users in India. All of their interactions would be pivotal to fine-tune the model, and the data flywheel we would build in the process would also give us a moat in the market. We also plan to perform some user-experience studies/interviews to ensure that we meet the doctors where they are and they can onboard with minimal friction.

18 Acknowledgments

I extend my heartfelt gratitude to Prof. Eric Nyberg, who has been an invaluable advisor throughout this project. His guidance and deep insights into the nuances of developing a healthcare tool have been instrumental in shaping Diagnolingo. Additionally, I would like to express my sincere thanks to my elder brother, Akshat Agarwal, for his unwavering support and for the numerous brainstorming sessions we've had. His perspective and ideas have significantly contributed to the development of this project, providing me with encouragement and creative solutions at every step.

19 Terminology, Definitions, Acronyms, and Abbreviations

Term	Description
SOAP	Subjective, Objective, Assessment, Plan
RAG	Retrieval Augmented Generation
EHR	Electronic Health Record
NER	Named Entity Recognition
LLM	Large-Language Model
SOTA	State-of-the-Art
API	Application Programming Interface
GUI	Graphical User Interface
NLP	Natural Language Processing
UI/UX	User Interface/Experience
OPD	Outpatient Department
Speech-to-Text (STT)	Conversion of spoken language into written text
Code-Switching	Alternating between two or more languages or dialects
Medical Lexicon	Vocabulary specific to medical terminology
Schema	Structure or format for organizing data
Transcription Accuracy	Degree of correctness in converting speech to text
Diarization	Speaker segmentation and labeling in audio recordings
Predictive Analytics	Analytical techniques for making predictions from data
Longitudinal EHR	EHR data collected over time for a single patient
Usability Testing	Evaluation of a system's ease of use by users
Workflow Integration	Seamless integration of system functionalities into existing workflows
Physician Engagement	Active involvement of doctors in the development process

20 References

- [1] Notification of Electronic Health Record Standards for India: <https://main.mohfw.gov.in/sites/default/files/17739294021483341357.pdf>
- [2] India launches incentive scheme to boost digital health record adoption: <https://www.healthcareitnews.com/news/asia/india-launches-incentive-scheme-boost-digital-health-record-adoption>
- [3] Electronic Health Record (EHR) Standards For India: https://www.nrce.in/download/files/pdf/nrce_ehr_stand_india.pdf
- [4] Need of the Hour: India needs to implement a nationwide framework for the adoption of EHRs: <https://timesofindia.indiatimes.com/blogs/voices/need-of-the-hour-india-needs-to-implement-a-nationwide-framework-for-the-adoption-of-ehrs/>
- [5] Yim, W.W., Fu, Y., Ben Abacha, A., Snider, N., Lin, T. and Yetisgen, M., 2023. Aci-bench: a novel ambient clinical intelligence dataset for benchmarking automatic visit note generation. *Scientific Data*, 10(1), p.586.
- [6] Abacha, A.B., Yim, W.W., Fan, Y. and Lin, T., 2023, May. An empirical study of clinical note generation from doctor-patient encounters. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 2291-2302).
- [7] Puri, N., Gupta, A., Aggarwal, A.K. and Kaushal, V., 2012. Outpatient satisfaction and quality of health care in North Indian medical institute. *International Journal of Health Care Quality Assurance*, 25(8), pp.682-697.
- [8] Kruse, G.R., Hays, H., Orav, E.J., Palan, M. and Sequist, T.D., 2017. Meaningful use of the indian health service electronic health record. *Health services research*, 52(4), pp.1349-1363.
- [9] Powell, A.C., Ludhar, J.K. and Ostrovsky, Y., 2017. Electronic health record use in an affluent region in India: Findings from a survey of Chandigarh hospitals. *International journal of medical informatics*, 103, pp.78-82.
- [10] Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C. and Sutskever, I., 2023, July. Robust speech recognition via large-scale weak supervision. In *International conference on machine learning* (pp. 28492-28518). PMLR.
- [11] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.T., Rocktäschel, T. and Riedel, S., 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33, pp.9459-9474.

A Appendix

The links given here serves as information useful for us in the capstone project.

A.1 Similar Speech-to-Text Products in USA

1. https://www.3m.com/3M/en_US/health-information-systems-us/create-time-to-care/clinician-solutions/speech-recognition/
2. <https://augmedix.com/products/>
3. <https://www.nuance.com/healthcare/case-study/temple-health.html>
4. https://www.praxisemr.com/speech_recognition.htm
5. <https://klasresearch.com/best-in-klas-ranking/speech-recognition-front-end-emr/2023/317>
6. <https://www.nextgen.com/solutions/provider-experience/speech-to-text>

A.1.1 Augmedix

Notebuilder app insight: <https://youtu.be/AG5HWyTODgw?si=evDw23LjXEixwDpi>

Suite of Products: <https://augmedix.com/products/>

Uses Google Cloud's healthcare solution, MedPalm2

609 **A.2 Indian Competition in broad Medical space**

- 610 1. <https://www.practo.com/>
- 611 2. <https://fastchart.com/>
- 612 3. <https://healthplix.com/>
- 613 4. <https://www.myhealthspace.in/for-doctors>
- 614 5. <https://www.linkedin.com/company/ekacare/>
- 615 6. <https://wellscribe.in/>
- 616 7. <https://www.f6s.com/company/karetakerabout>
- 617 8. <https://www.f6s.com/company/onito>
- 618 9. <https://www.f6s.com/company/thedocwebabout>
- 619 10. <https://www.linkedin.com/company/fololife/>

620 **A.3 Speech-to-Text dataset Ideas**

- 621 1. <https://web.stanford.edu/class/cs224s/datasets/>
- 622 2. <https://mlcommons.org/datasets/peoples-speech/>
- 623 3. <https://paperswithcode.com/datasets?task=speech-recognition>
- 624 4. BioAsk Q&A Datasets
- 625 5. TedX Medical Talks

626 **A.4 Vernacular dataset sources**

- 627 1. <https://www.openslr.org/103>
- 628 2. <https://www.twine.net/blog/top-indian-language-datasets/>
- 629 3. <https://npl.t.in/demo/speech-corpus>
- 630 4. <https://www.twine.net/blog/hindi-language-datasets/>
- 631 5. <https://www.shaip.com/offering/speech-data-catalog/hindi-dataset/>
- 632 6. <https://paperswithcode.com/datasets?mod=audio&lang=hindi>
- 633 7. <https://www.datatang.ai/datasets/1156>
- 634 8. <https://www.futurebeeai.com/dataset/speech-data/hindi-dataset>
- 635 9. <https://www.surfing.ai/services/recognition/>
- 636 10. <https://defined.ai/>

637 **A.5 Incumbent APIs**

638 **A.5.1 Google Cloud's NER**

639 NER etc APIs: https://youtu.be/pUnuLkEedfE?si=yxTv_1ZyBUdyxybw

640 Documentation: <https://cloud.google.com/healthcare-api/docs/concepts/nlp>

641 Pricing: <https://cloud.google.com/healthcare-api/pricinghealthcare-natural-language-api> Pricing is

642 roughly Rs 5-10 per minute's worth of text (Entity Recognition for 1 Text Record is \$0.10, 1 Text

643 Record is 1000 characters, 1000 characters is 250 words, 250 words is spoken in 1-2 minutes).

644 **A.5.2 OpenAI Whisper**

645 Demo: <https://huggingface.co/spaces/openai/whisper>

646 Can prompt model (240 characters) for added context.

647 Pricing: <https://openai.com/pricing#audio-models> (Rs 1/min)

648 Not Streaming, Batches of 30s Diarization? Code Switching? - kind of possible - might not be

649 accurate No Medical Lexicons

650 **A.5.3 Google STT**

651 Has medical transcription support (just for English) (Pricing is roughly Rs 7.5 per minute) Google
652 STT - Hindi, Kannada, Malayalam, Marathi, Tamil, Telugu, etc. available for long/short formats -
653 also some other Indian languages supported by default only API STT v2 is latest (Chirp) No support
654 for Diarization for Indian Languages No Code Switching

655 **A.5.4 Amazon Transcribe**

656 Has medical transcription support (just for English) (Pricing is roughly Rs 7.5 per minute) Has
657 Hindi, etc Indian languages supported (but mostly batching) Need to check for Diarization No Code
658 Switching

659 **A.6 Code & Open-Source Models**

660 Open Source LLMs: Meta (free for commercial use): <https://llama.meta.com/>

661 Code Samples: [https://www.kaggle.com/code/quadeer15sh/unsupervised-learning-on-hindi-speech-](https://www.kaggle.com/code/quadeer15sh/unsupervised-learning-on-hindi-speech-audio-files)
662 [audio-files](https://www.kaggle.com/code/quadeer15sh/unsupervised-learning-on-hindi-speech-audio-files)

663 <https://huggingface.co/blog/audio-datasets>

664 Medical Text Generation Code Sample: <https://github.com/Yale-LILY/Ascle>

665 **B Changes To Previous Deliverables**

666 Initially, the idea was to first implement a speech-to-text pipeline alongside data collection. However,
667 due to delays in setting up data sharing standards with doctors, decided to move ahead with text-to-
668 EHR first as there are publicly available doctor-patient conversation transcripts in English. Then
669 we returned to the first part of the pipeline by mocking the data we need. We also leveraged
670 state-of-the-art Whisper models by OpenAI instead of developing a Speech-to-Text module from
671 scratch.

672 **C Feedback on Draft Report**

673 Initially my results section was not complete and I lost marks there. Therefore I updated the same
674 and resubmitted with the final results and received full marks for the draft report.

675 **D Feedback during Capstone Presentation**

676 None