**Perclias: Personal Clinical Assistant**

**Project Overview:**

**Scope:**

Perclias is a conversational AI assistant designed to assist patients and healthcare professionals in managing clinical information and providing personalized healthcare solutions. The project will leverage natural language processing (NLP) to capture and analyze patient-doctor conversations, generate summaries, and provide tailored recommendations based on the identified symptoms and medical history.

The project will involve the following key components:

* Voice Recording and Transcription: Perclias will record and transcribe patient-doctor conversations using speech recognition technology.
* Conversation Summarization: The transcribed conversations will be processed using NLP techniques to generate concise summaries, highlighting key points and medical information.
* Symptom Identification and Analysis: The summaries will be analyzed to identify relevant symptoms, medical conditions, and potential diagnoses.
* Recommendation Generation: Based on the identified symptoms and medical history, Perclias will provide personalized recommendations, treatment options, and self-care advice.
* Report Generation and Communication: Perclias will generate comprehensive reports summarizing the conversation, identified symptoms, and recommended actions. These reports can be edited, saved as PDFs, and shared with patients via email.

**Stakeholders:**

The primary stakeholders for the Perclias project include:

**Patients:** Individuals seeking medical advice and personalized healthcare solutions.

**Healthcare Professionals:** Doctors, nurses, and other medical practitioners who interact with patients and require assistance in managing clinical information and providing recommendations.

**Healthcare Organizations:** Hospitals, clinics, and other healthcare facilities that can benefit from streamlined patient-doctor communication and improved patient care.

**Problem Statement:**

**Current Challenges:**

* Inefficient Information Management: Manual capture and organization of patient-doctor conversations can be time-consuming and prone to errors, leading to inefficiencies in healthcare delivery.
* Limited Personalization: Providing personalized healthcare recommendations based on individual patient symptoms and medical history is challenging, particularly in high-volume healthcare settings where time constraints are prevalent.
* Communication Barriers: Effective communication between patients and healthcare professionals is essential for accurate diagnosis and treatment. However, language barriers, medical jargon, and limited consultation time often impede this process, hindering the quality of care provided.

**Opportunities:**

* Enhanced Patient Care: By streamlining the capture and analysis of patient-doctor conversations, Perclias can assist healthcare professionals in providing more personalized and accurate care, ultimately leading to better patient outcomes.
* Increased Efficiency: Automating the process of summarizing conversations and generating recommendations can save valuable time for healthcare professionals, allowing them to allocate their resources more effectively and focus on critical tasks.
* Improved Accessibility: Perclias can empower patients by providing clear and understandable summaries of their medical information, facilitating better self-care and informed decision-making. Additionally, it can help overcome communication barriers by providing multilingual support and simplifying medical terminology.

**Methodology:**

**Data Sources:**

Perclias will primarily rely on patient-doctor conversations as the primary data source. These conversations can be recorded and transcribed using speech recognition technology or synthesized using natural language generation techniques for testing and development purposes.

**Technologies and Tools**

The following technologies and tools will be utilized in the development of Perclias:

* Natural Language Processing (NLP): Libraries such as spaCy, NLTK, or Hugging Face Transformers will be used for text processing, summarization, and symptom identification.
* Speech Recognition: Libraries like Google Speech-to-Text, Mozilla DeepSpeech, or OpenAI Audio API(Whisper Model) will be used for voice recording and transcription.
* Web Development: Streamlit, FastAPI will be used for building the user interface and deploying the application.
* Databases:Pinecone and Snowflake or other suitable databases will be used for storing conversation transcripts, summaries, and patient information.
* Cloud Services: Cloud platforms like AWS, Google Cloud, or Microsoft Azure may be utilized for scalable deployment, data storage, and processing.

**Pipeline Design:**

The Perclias pipeline will consist of the following stages:

* Data Ingestion: Patient-doctor conversations will be recorded and transcribed using speech recognition technology or synthesized for testing purposes.
* Data Preprocessing: Transcribed conversations will undergo text cleaning, normalization, and tokenization to prepare the data for further processing.
* Conversation Summarization: NLP techniques, such as extractive or abstractive summarization, will be applied to generate concise summaries of the conversations.
* Symptom Identification and Analysis: The summaries will be analyzed using LLMs to identify relevant symptoms, medical conditions, and potential diagnoses.
* Recommendation Generation: Based on the identified symptoms and medical history, personalized recommendations, treatment options, and self-care advice will be generated using rule-based systems.
* Report Generation: Comprehensive reports will be generated, summarizing the conversation, identified symptoms, and recommended actions.
* User Interface: A user-friendly interface will be developed using web development frameworks like Streamlit or FastAPI, allowing users to interact with Perclias, review and edit summaries and recommendations, and generate reports.
* Communication and Storage: The generated reports can be saved as PDFs, shared with patients via email, and stored in a database for future reference.

**Data Processing and Transformation:**

The data processing and transformation stage will involve the following steps:

* Text Cleaning: Removing irrelevant information, such as stop words, punctuation, and special characters, from the transcribed conversations.
* Normalization: Standardizing text by converting to lowercase, expanding contractions, and handling misspellings.
* Tokenization: Breaking the text into smaller units, such as words or sentences, for further processing.
* Named Entity Recognition (NER): Identifying and extracting relevant medical entities, such as symptoms, conditions, and treatments, from the text.
* Symptom Extraction: Applying NLP techniques and LLMs models to identify and classify symptoms based on the extracted medical entities.
* Recommendation Generation: Utilizing rule-based systems to generate personalized recommendations based on the identified symptoms and medical history.

**Data Sources:**

1. WHO Global Health Observatory (GHO) resources by the World Health Organization: This includes data sets and reports from 194 countries on a wide variety of health topics.

2. DHS Program Medical datasets: These datasets span multiple health topics and are based on surveys, biomarker testing, and geographic data from around the globe.

3. HealthData.gov: The official US government healthcare website, which includes multiple open datasets on topics ranging from COVID-19 to health equity.

4. Life Science Database Archive: A life science dataset from Japan, gathered by life scientists, including datasets about organs, antigens, chemicals, and more.

5. Data.gov.au: The official source of Australian open government data, which includes healthcare datasets.

6. Kent Ridge Biomedical Datasets: Datasets from the biomedical field, including journal-published data.

7. CDC WONDER: A US CDC database with public health information around topics like mortality, natality, cancer, and vaccinations.

8. OpenFDA: APIs and raw download access to structured FDA datasets on adverse events, drug product labeling, and recall enforcement reports.

9. The Big Cities Health Inventory Data Platform: Over 150,000 data points from 35 large US cities across 120 health-related metrics.

10. OECD Health Statistics: Comparable statistics on health and health systems across OECD countries.