

## Medical Report Generation GenAI

Generate clinical notes, discharge summaries, and medical transcriptions from patient data using AI.

### Project Overview

MedicalReportGenAI is an AI-powered system designed to automate the generation of clinical notes and discharge summaries from patient data. It leverages state-of-the-art NLP models (Transformers, GPT-2) and ML pipelines to process patient records, extract relevant symptoms, and produce structured clinical notes, discharge summaries, and medical transcriptions.

This tool helps medical teams save time, improve documentation accuracy, and generate actionable patient information efficiently.

### Business Value

Healthcare providers and data teams can use this system to:

- Automatically generate structured clinical notes from patient data.
- Produce discharge summaries for faster patient discharge processes.
- Transcribe doctor dictations for medical transcription.
- Reduce manual documentation effort and improve record-keeping accuracy.
- Maintain logs and metrics for analytics and auditing using MLflow.

### Technical Workflow

The solution integrates multiple AI modules to process patient data:

1. Data Loading – Load prescription and patient data from CSV files.
2. NER Extraction – Extract symptoms and relevant entities using a clinical NER model (bert-base-uncased\_clinical-ner).
3. Clinical Note Generation – Generate structured clinical notes using GPT-2 based text-generation.
4. Discharge Summary Generation – Produce discharge summaries with patient history, medications, and extracted symptoms.
5. Medical Transcription – Transcribe doctor audio dictations using whisper.

6. MLflow Logging – Track metrics, parameters, and artifacts for reproducibility.
7. Streamlit UI – Interactive dashboard to input patient symptoms and view generated outputs.

### Tech Stack

- Python 3.9+ – Core programming language
- Transformers – NLP models for text generation and entity extraction
- Whisper – Audio transcription
- MLflow – Experiment tracking and logging
- Streamlit – Interactive frontend UI
- Docker – Containerized deployment
- Kubernetes – Scalable deployment configuration
- Pandas – Data processing and analysis

### Project Structure

```
|—— K8/           # Kubernetes deployment and service configs  
|—— Transformers/    # Pretrained models and checkpoints  
|—— dataset/        # Generated patient records  
|—— hosp/          # Source patient prescription CSV files  
|—— mlruns/         # MLflow logs  
|—— .idea/  
|—— .vscode/  
|—— __pycache__/  
|—— .dockerignore  
|—— .gitignore  
|—— Dockerfile       # Docker image build file  
|—— clinical_note_generator.py # Generate clinical notes from patient data  
|—— data_loader.py      # Load prescription and patient data
```

```
|-- discharge_summary.py      # Generate discharge summaries  
|-- main.py                  # Streamlit UI application  
|-- medical_transcription.py # Audio-to-text medical transcription  
|-- ner_entity.py            # Clinical Named Entity Recognition  
└-- requirements.txt         # Python dependencies
```

### How to Run Locally

1. Clone the repository:

```
git clone https://github.com/Prashanth-practise-account/Medical-Report-Generation  
cd GenAI_Clinical_Note_Generation
```

2. Install dependencies:

```
pip install -r requirements.txt
```

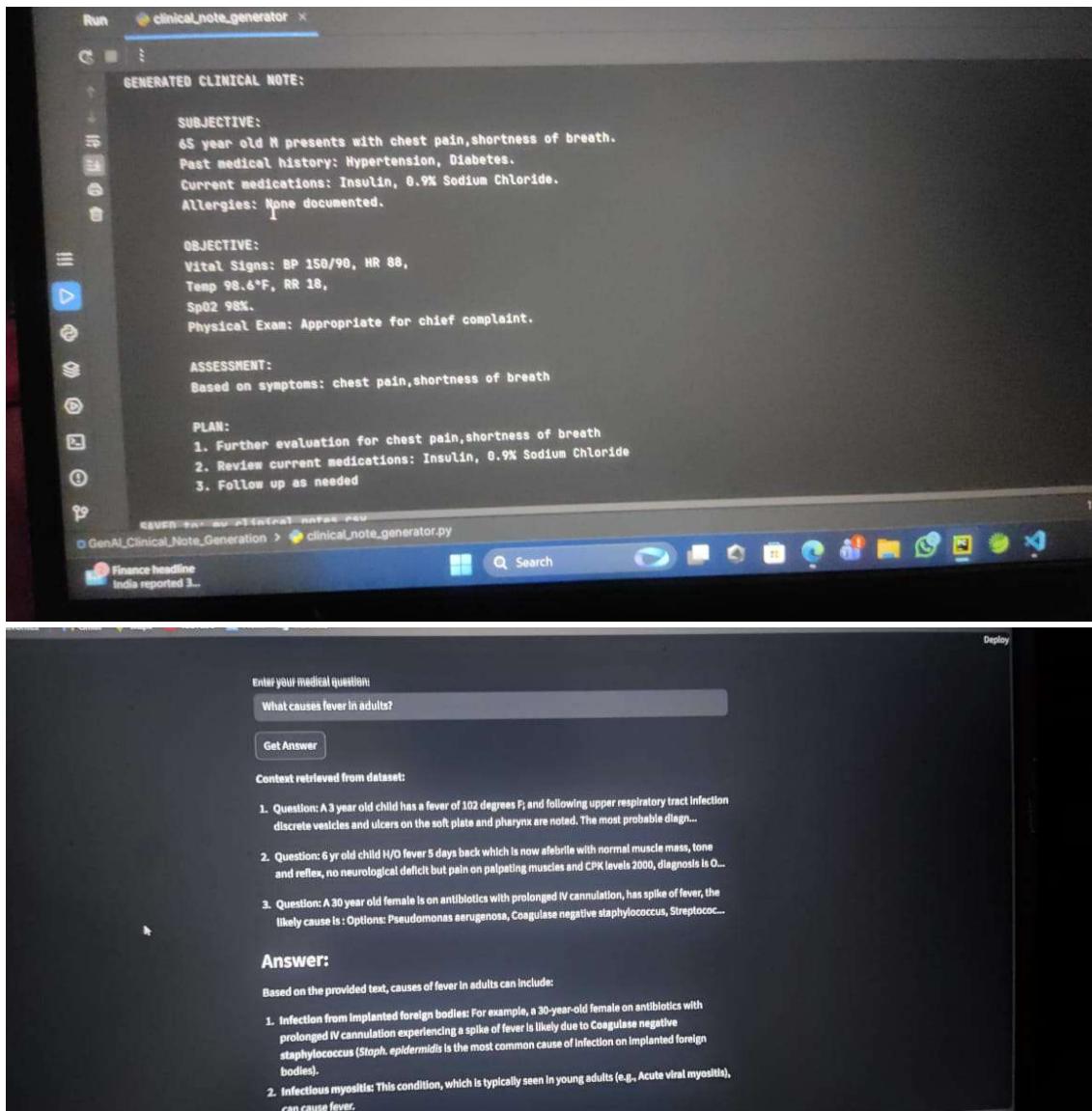
3. Run the Streamlit app:

```
streamlit run main.py
```

4. Use the app in your browser to input patient symptoms, generate clinical notes, discharge summaries, and medical transcription outputs.

### How to Use

1. Input Patient Symptoms – Enter symptoms in the Streamlit UI.
2. Generate Clinical Note – Extract symptoms, medications, and history to create a structured clinical note.
3. Generate Discharge Summary – Summarize patient data, medications, and recommendations.
4. Medical Transcription – Upload audio recordings to transcribe doctor dictations.
5. Save Outputs – Generated notes and summaries are saved to CSV (dataset/my\_patient\_records.csv) and logged in MLflow.



## 💡 Customization Ideas and Future Enhancements

- Vectorized Clinical Data Analysis – Store generated notes in vector databases for faster semantic search.
- Real-time Voice Transcription – Support live dictation during patient consultations.
- Multi-Language Support – Extend NLP models for multiple languages.
- Integration with EHR Systems – Automate note generation directly from Electronic Health Records.
- Enhanced Security & Compliance – HIPAA/GDPR compliant storage and logging.

 Author

Prashanth B H – AI/ML Developer | Streamlit & Transformers Enthusiast

- GitHub: <https://github.com/Prashanth-practise-account>
- Contact: bhprashanth0@gmail.com