GEN-AI HEALTHCARE AGENT

Team-1 (Members):

Hamsalakshmi Ramachandran - 017423666

Himani Shah - 017411407

Kush Bindal - 017441359

Shobhita Agrawal - 017552795

Sugandha Chauhan - 017506190

INTRODUCTION

GenMedX is an innovative healthcare project that transforms emergency department data into an intelligent AI agent system. Using a sophisticated cloud pipeline built with Apache Airflow, the project processes MIMIC-IV ED clinical data from PhysioNet through automated cleaning and transformation workflows. For our Named Entity Recognition module, we utilize the NCBI Disease Corpus.

PROJECT BACKGROUND

- Healthcare today is strained by an aging population, rising costs, and workforce shortages.
- Care gaps emerge when patients leave clinical settings, leading to oversight lapses, delayed interventions, and suboptimal resource use.
- Studies show 30 % of sudden complications are preventable with better education, and ED readmissions alone exceed $17 billion annually.

EXECUTIVE SUMMARY

GenMedX aims to deliver a suite of Generative AI agents spanning the emergency-care continuum:

- Clinical Explanation Agent: Transforms complex diagnoses into clear, patient-friendly explanations.
- ED Stay Summarizer Agent: Generates concise, chronologically ordered narratives of each ED encounter.
- Post-ED Follow-Up Agent: Provides personalized discharge plans, appointment scheduling, and readmission-risk monitoring.

PROJECT REQUIREMENTS

- Data Ingestion & Transformation
- Fine-Tuning
- RAG
- Evaluation & Quality Assurance
- Performance, Resource & Privacy Constraints

TEAM ORGANIZATION AND FUNCTION ROLES

Team Member	Responsibilities
Hamsa	ETL pipeline, Airflow implementation, Report writing
Sugandha	BigQuery datapipeline implementation, Report Writing
Himani	Build RAG pipeline for clinical explanation agent, Report Writing
Kush	NER finetuning approach for transfer learning, Report Writing
Shobhita	NER prompt engineering approach for transfer

PROJECT LITERATURE SURVEY

Recent studies demonstrate that LLMs can substantially streamline emergency department documentation and patient education when integrated with clinician oversight.

→ **Hartman et al. (2024)** showed a tailored clinical model markedly outperformed physician-written handoff notes on key NLP metrics, though clinicians rated AI summaries as slightly less useful and safe.

→ **De Rouck et al. (2025)** found GPT-4–generated discharge brochures scored well on quality and clarity but required human editing to meet recommended reading levels.

→ A systematic review by **Meyer and Meyer (2025)** further confirms that ChatGPT-4 enhances administrative workflows—such as summaries and handoffs—while underscoring heterogeneity in study methods and the critical need for clinician supervision.

We are planning to compare our LLM-generated handoff notes and summaries against physician-written ones using ROUGE, BERTScore, and inconsistency metrics, and survey clinicians on usefulness and safety with the same scales to measure any improvements.

Hartman, V., Zhang, X., Poddar, R., McCarty, M., Fortenko, A., Sholle, E., ... & Steel, P. A. (2024). Developing and evaluating large language model–generated emergency medicine handoff notes. *JAMA Network Open*, *7*(12), e2448723-e2448723. https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2827327

De Rouck, R., Wille, E., Gilbert, A., & Vermeersch, N. (2025). Assessing artificial intelligence-generated patient discharge information for the emergency department: a pilot study. *International Journal of Emergency Medicine*, *18*(1), 85. https://link.springer.com/article/10.1186/s12245-025-00885-5

Meyer, N. S., & Meyer, J. W. (2025). A Practical Guide to the Utilization of ChatGPT in the Emergency Department: A Systematic Review of Current Applications, Future Directions, and Limitations. *Cureus*, *17*(4). https://assets.cureus.com/uploads/review_article/pdf/346933/20250506-257604-8ki60b.pdf

PROJECT REQUIRED RESOURCES, TECHNOLOGY AND PLATFORM

Hardware Resources

University GPU Workstation – NVIDIA GPU (16GB VRAM)
Google Colab Pro – On-demand GPU (T4/P100)
Cloud Storage – Google Cloud Platform & Drive

Software & Tools

ML Frameworks – PyTorch, TensorFlow, Hugging Face Transformers
Pretrained Models – BioClinicalBERT (fine-tuned on MIMIC-IV & NCBI)
IDE & Dev Tools – Jupyter Notebook, VS Code, GitHub (version control)
BigQuery

Project Management & Visualization

Gantt chart and PERT chart – Task tracking & scheduling

MODEL SELECTION, INNOVATION AND COMPARISON AND JUSTIFICATIONS

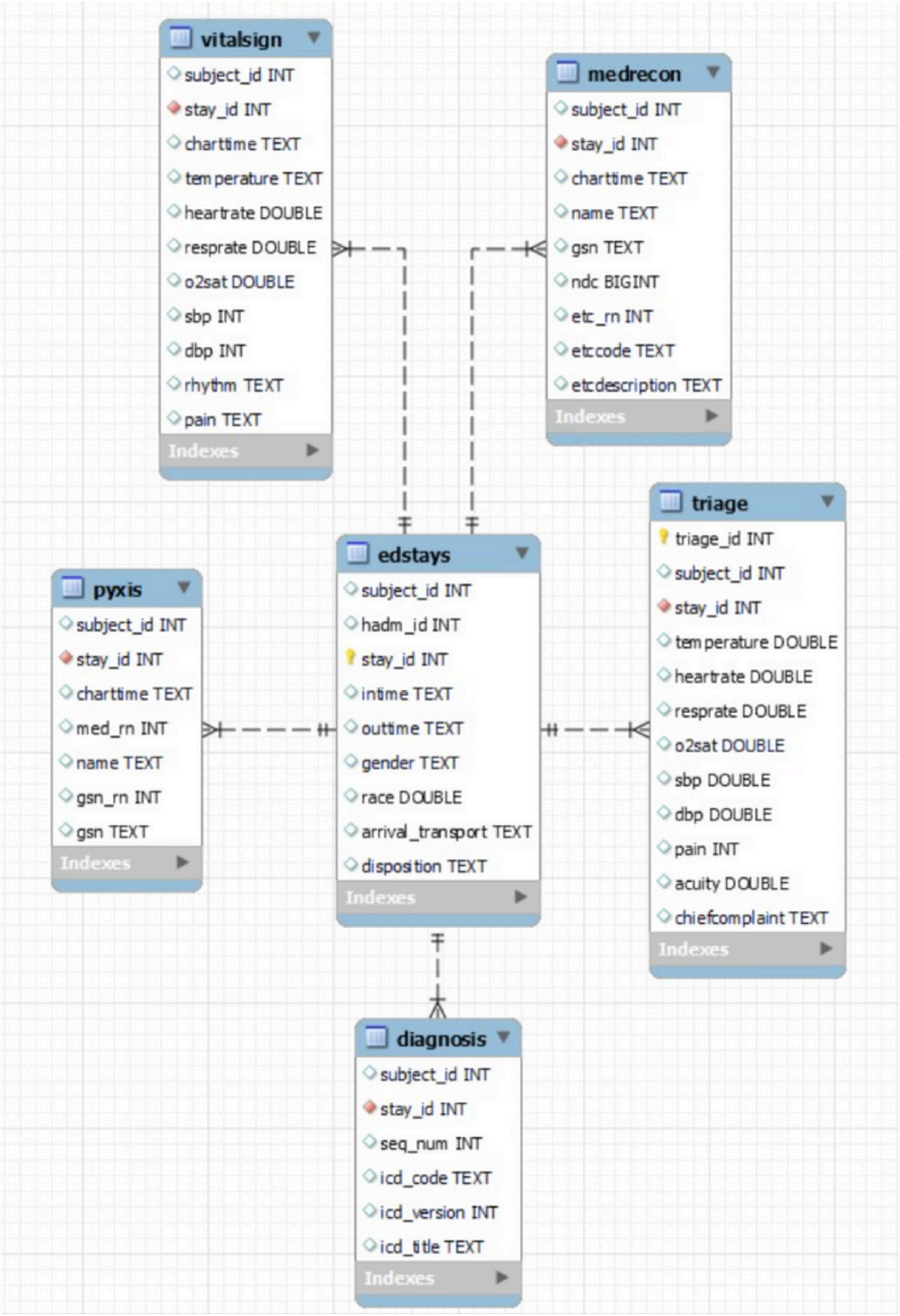
Model	Strengths	Limitations	Use in GenMedX	Techniques
MedAlpaca-7B	Strong medical domain knowledge, excellent at clinical text generation	Computationally demanding; may require GPU resources; can occasionally generate inaccuracies without precise contexts	Primary model for Clinical Explanation Agent; Clinical summarization through transfer learning from PubMed to ED dataset	LoRA fine-tuning; RAG integration; Transfer learning from PubMed data
BioMistral 7B	Biomedical fine-tuning, efficient for summarization	Still requires LoRA fine-tuning for task-specific adaptation	Primary model for summarizer	LoRA (Low-Rank Adaptation); RAG Generator; Extended context window for longitudinal summaries
Meditron (LLaMA 2)	Medical focus, good base for summarization	May require additional fine-tuning for ED-specific nuances	Backup or ensemble option	LoRA or QLoRA fine-tuning; can serve as backup summarizer; works with RAG
PatientSeek	Long-context handling, customizable	No inherent medical knowledge, requires intense fine-tuning	Considered for longitudinal summaries	PEFT + Long-context RAG; context window expansion; can power patient history retrieval
OpenBioLLM	Very large capacity, strong medical reasoning	High computational cost, difficult real-time deployment	Research exploration	Full fine-tuning only; supports multi-hop reasoning; can power few-shot learning on demand
BioGPT / PubMedGPT	Efficient, good at biomedical explanations	Not powerful enough for full summarization	Supplementary for patient-friendly summaries	Lightweight generation; used for patient-friendly output modules; zero-shot answers

DATASET-1

The **MIMIC-IV-ED** (v2.2) dataset is a large-scale, publicly available resource that contains detailed records of approximately 425,000 emergency department (ED) visits that took place at the Beth Israel Deaconess Medical Center (BIDMC) between 2011 and 2019. It is part of the broader MIMIC-IV data ecosystem and is hosted on PhysioNet, a platform for freely sharing medical data for research.

BIG DATA STATISTICS

Table Name	Description	Record Count
edstays	Emergency department visits (unique stays IDs)	~425,087
triage	Triage assessment, vitalsigns, chief complaints	~425,087
vitalsign	Vital Signs recorded throughout the ED stay	~1,564,610
diagnosis	Discharge diagnosis with ICD-10 codes	~899,050
medrecon	Medication reconciliation at admission	~2,987,342
pyxis	Administered medications during ED stays	~1,586,053



STRUCTURED DATA:

- PATIENT DEMOGRAPHICS (AGE, GENDER, ETHNICITY)
- EMERGENCY DEPARTMENT (ED) ENCOUNTERS (ARRIVAL TIME, TRIAGE INFORMATION, DISPOSITION)
- DIAGNOSIS CODES (ICD-9 AND ICD-10)
- PROCEDURE CODES (CPT, HCPCS)
- LABORATORY TEST RESULTS (CHEMISTRY, HEMATOLOGY, MICROBIOLOGY)
- VITAL SIGNS (HEART RATE, BLOOD PRESSURE, TEMPERATURE, RESPIRATORY RATE, AND OXYGEN SATURATION) WERE RECORDED AT VARIOUS INTERVALS.
- MEDICATION ADMINISTRATION RECORDS DURING THE ED STAY.
- DATA RELATED TO MEDICAL DEVICES AND INTERVENTIONS USED IN THE ED.

SEMI-STRUCTURED DATA

- TRIAGE NOTES DETAILING THE INITIAL ASSESSMENT AND CHIEF COMPLAINTS.
- PROVIDER NOTES DOCUMENTING THE PATIENT'S CONDITION, TREATMENT PLAN, AND PROGRESS.
- DISCHARGE SUMMARIES OUTLINING THE ED COURSE AND RECOMMENDATIONS

DATASET-2

NCBI is a collection of 5000 rows and 793 PubMed abstracts from biomedical research literature discussing various medical topic.

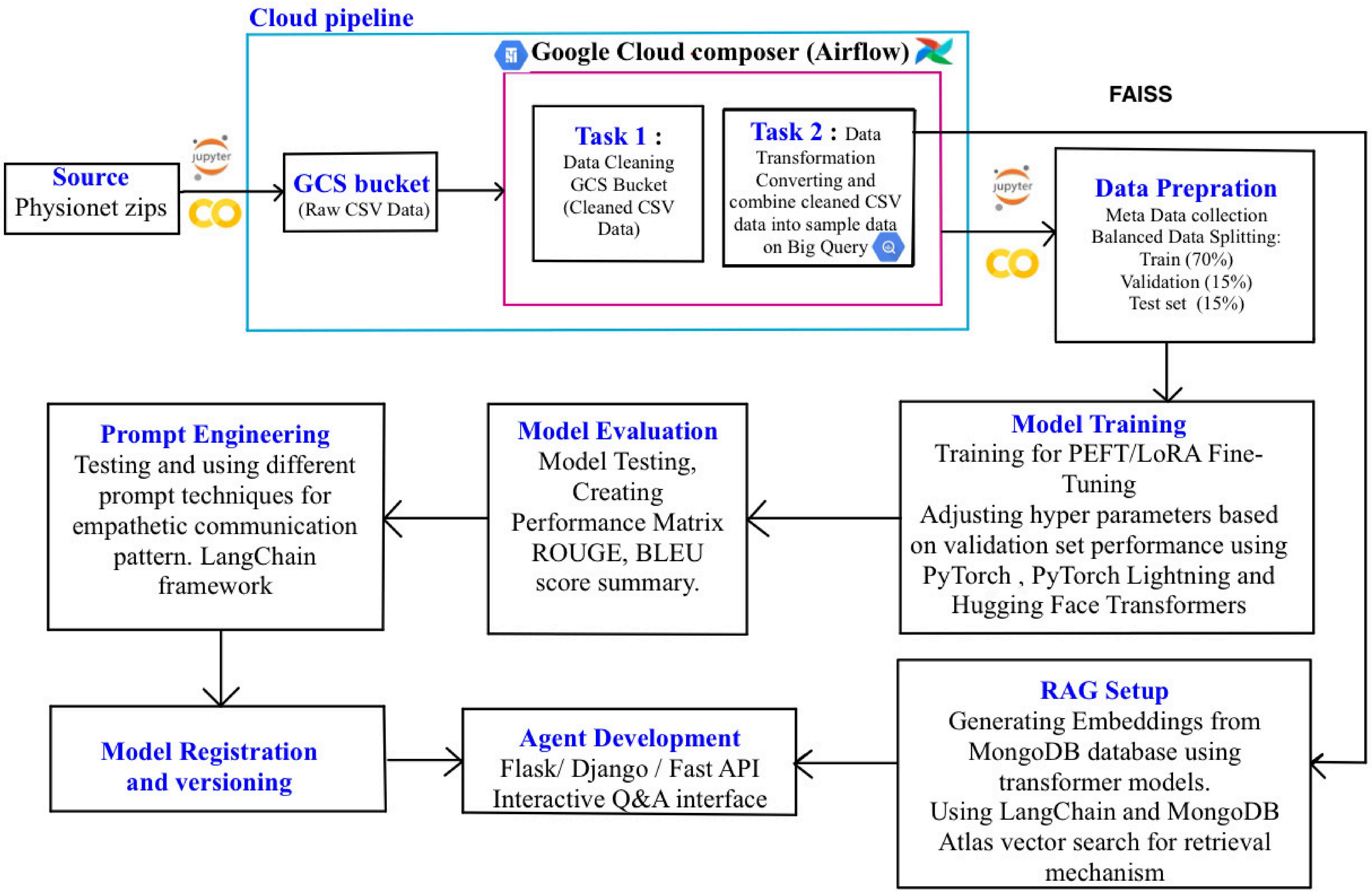
Publicly available with complex medical terminology suitable for both fine-tuning and LLM evaluation.

For annotations, there is expert-marked disease name recognition using BIO format (B-Disease, I-Disease, O)

Example 1:

Token	Tag ID	NER Tag
Identification	0	O
of	0	O
APC2	0	O
,	0	O
a	0	O
homologue	0	O
of	0	O
the	0	O
adenomatous	1	B-Disease
polyposis	2	I-Disease
coli	2	I-Disease
tumour	2	I-Disease
suppressor	0	O
.	0	O

PROJECT DEVELOPMENT METHODOLOGY

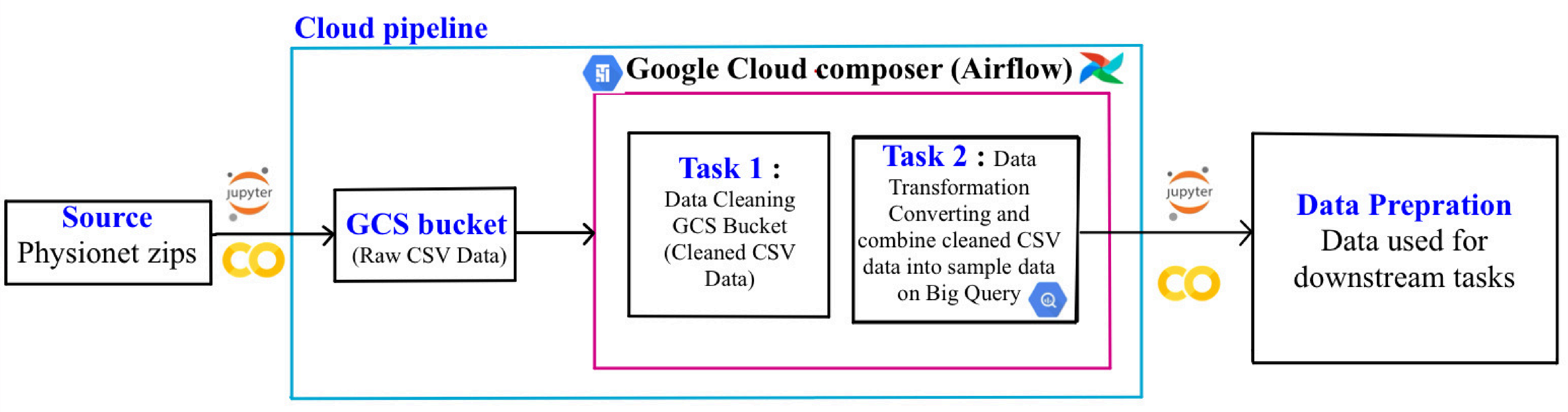


DATA PIPELINE

AUTOMATED CLOUD PIPELINE

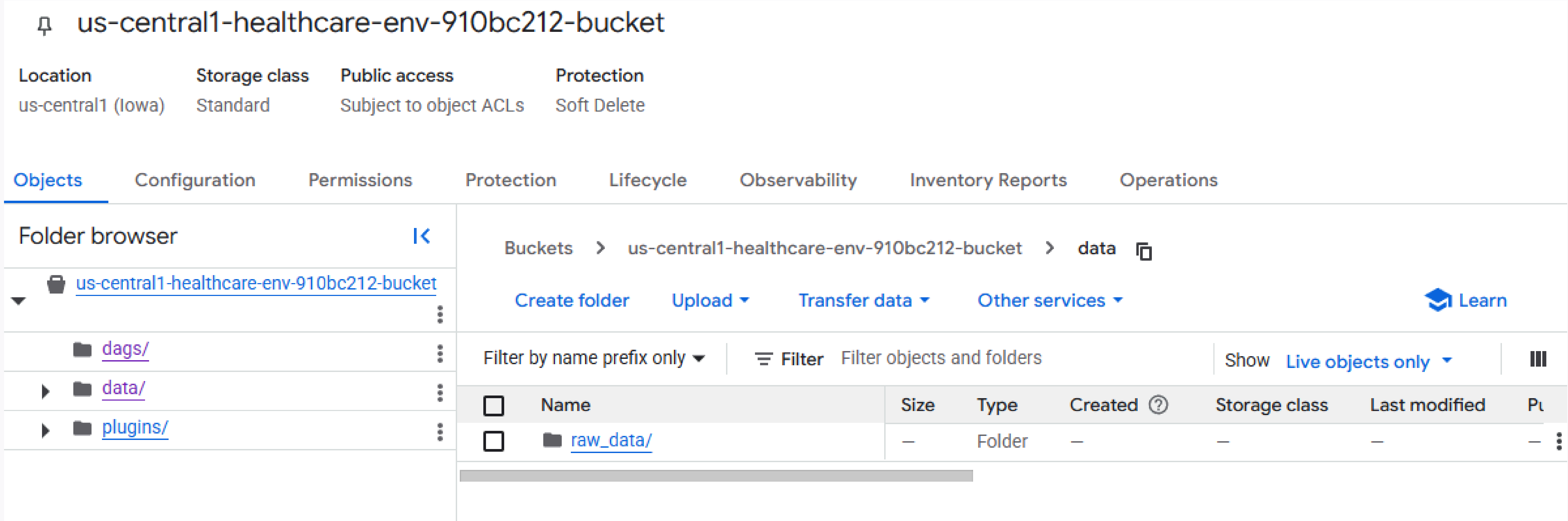
Sucessfully processed aproximately 425,000 ER visits and established a structured cloud-based ETL pipeline using Google Cloud Platform

- Google Cloud Storage
- Google Cloud Composer
- Google Big Query



DATA PIPELINE

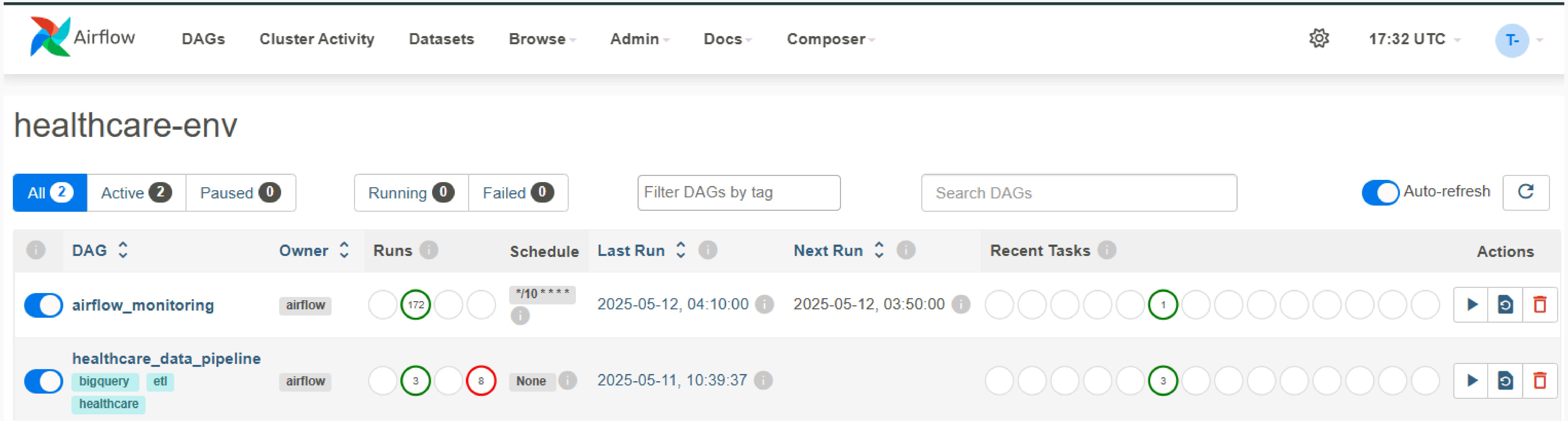
- **Ingestion-** Raw PhysioNet zips decompressed via Jupyter/Colab and uploaded to GCS bucket under healthcare directory.



DATA PIPELINE

Cloud-Orchestrated Infrastructure:

- Google Cloud Composer (managed Apache Airflow)
- ETL DAG (healthcare_data_pipeline)
- Ensures modular, scalabel, and reproducible data workflows



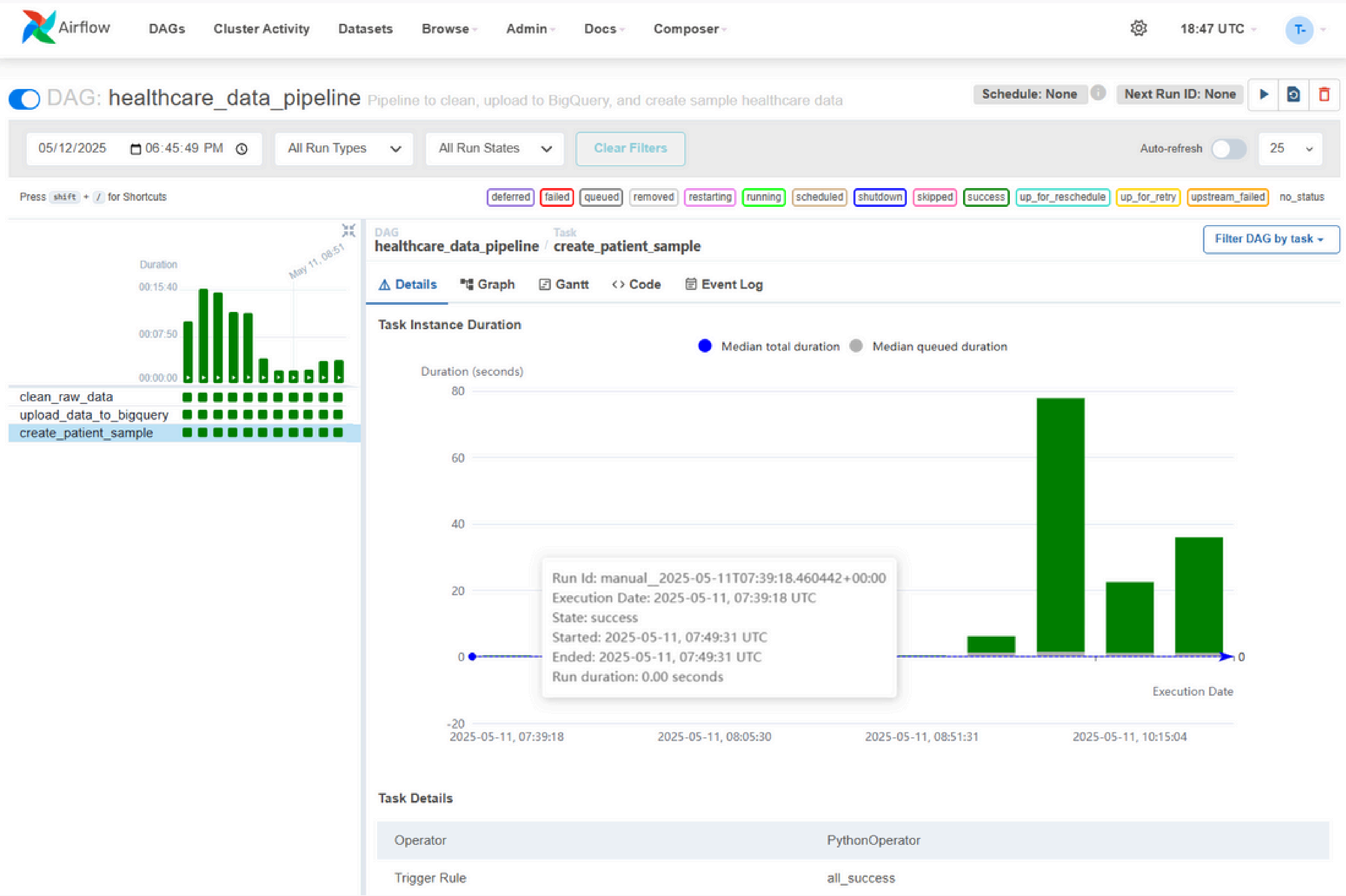
DATA PIPELINE

Pipeline monitoring

- Real-time tracking of task execution using the Airflow UI, with color-coded indicators (e.g., green for success, red for failure) for quick status assessment.

- Task Instance Duration charts help monitor performance trends and detect bottlenecks or anomalies across multiple DAG runs.

- Detailed metadata, logs, and timing info (start/end time, duration, state) available for each task instance, enabling precise debugging and pipeline optimization.



DATA PIPELINE

- Data Cleaning - Remove nulls, handling missing values and duplicates , normalizing clinical units and timestamps.
- Transformation- Merge & convert Cleaned CSVs and load them to BigQuery

sample_data_... Query Open in Share Copy Snapshot Delete Export

Schema Details Preview Table Explorer Preview Insights Lineage Data Profile Data Quality

Row	subject_id	hadm_id	stay_id	intime	outtime	gender	race
1	10021487	28998349.0	38319705	2116-12-02 22:57:00 UTC	2116-12-03 01:02:00 UTC	M	WHITE
2	10019568	28710730.0	36381328	2120-01-30 19:44:00 UTC	2120-01-30 22:51:00 UTC	F	WHITE
3	10021487	28998349.0	38319705	2116-12-02 22:57:00 UTC	2116-12-03 01:02:00 UTC	M	WHITE
4	10021487	28998349.0	38319705	2116-12-02 22:57:00 UTC	2116-12-03 01:02:00 UTC	M	WHITE
5	10004235	24181354.0	38926302	2196-02-24 12:15:00 UTC	2196-02-24 17:07:00 UTC	M	MULTIPLE RACE/ETHN
6	10021487	28998349.0	38319705	2116-12-02 22:57:00 UTC	2116-12-03 01:02:00 UTC	M	WHITE
7	10022620	27180902.0	34614222	2174-01-01 15:34:00 UTC	2174-01-01 18:52:05 UTC	M	UNKNOWN
8	10004235	24181354.0	38926302	2196-02-24 12:15:00 UTC	2196-02-24 17:07:00 UTC	M	MULTIPLE RACE/ETHN
9	10004235	24181354.0	38926302	2196-02-24 12:15:00 UTC	2196-02-24 17:07:00 UTC	M	MULTIPLE RACE/ETHN
10	10021487	28998349.0	38319705	2116-12-02 22:57:00 UTC	2116-12-03 01:02:00 UTC	M	WHITE
11	10004235	24181354.0	38926302	2196-02-24 12:15:00 UTC	2196-02-24 17:07:00 UTC	M	MULTIPLE RACE/ETHN
12	10021487	28998349.0	38319705	2116-12-02 22:57:00 UTC	2116-12-03 01:02:00 UTC	M	WHITE
13	10021487	28998349.0	38319705	2116-12-02 22:57:00 UTC	2116-12-03 01:02:00 UTC	M	WHITE
14	10021487	28998349.0	38319705	2116-12-02 22:57:00 UTC	2116-12-03 01:02:00 UTC	M	WHITE
15	10021487	28998349.0	38319705	2116-12-02 22:57:00 UTC	2116-12-03 01:02:00 UTC	M	WHITE

- PROPOSED MODEL AND FRAMEWORK-

AGENT 1 - CLINICAL EXPLANATION AGENT

Key Points:

- Provides accurate, personalized, and easy-to-understand clinical explanations.
- Converts complex patient data into clear, patient-centered narratives.
- Helps healthcare providers and patients understand medical information efficiently.

Framework Overview

- Data Extraction:
  - Collects patient details (diagnoses, medications, vital signs) from BigQuery.
- Semantic Embedding & Retrieval:
  - Matches patient queries to relevant clinical contexts using semantic embeddings and FAISS retrieval.
- Generation of Explanations:
  - Creates patient-specific clinical narratives using a specialized medical language model.

Model Details and Rationale:

- Sentence-Transformers (all-MiniLM-L6-v2) efficiently generates semantic embeddings, enabling precise, real-time retrieval of patient contexts crucial for clinical workflows.
- FAISS (faiss.IndexFlatL2) quickly searches high-dimensional embeddings, supporting rapid and scalable retrieval operations necessary for real-time applications.
- MedAlpaca-7B (Medical LLM) is specifically trained on medical data, reliably producing clinically accurate explanations and minimizing inaccuracies or unsupported claims.

Result:

```
Agent Response:
 You are a clinical decision support assistant.

Step 1: Analyze the patient's symptoms, vitals, and diagnosis.

Step 2: Consider whether the prescribed medication fits the case.

Step 3: Justify the decision medically.

Chief Complaint: PICC LINE EVAL
Vitals → Temp: nan°F, HR: nan bpm, RR: nan, BP: nan/nan, O2: nan%
Pain: 5, Acuity: 3.0
Demographics: M, WHITE, Transport: WALK IN
Diagnosis: MECH COMPL OF INFUSION CATHETER, INITIAL ENCOUNTER (T82594A)
Medication: VANCOMYCIN | ETC: ANALGESIC OR ANTIPYRETIC NON-OPIOID
Additional Med (Pyxis): OXYCODONE (IMMEDIATE REL 5MG TAB @ 2187-05-11 15:00:00+00:00
Vitals Time: 2187-05-11 19:25:00+00:00, HR: 73.0
Chief Complaint: LINE EVAL
Vitals → Temp: 98°F, HR: nan bpm, RR: nan, BP: 108/90, O2: nan%
Pain: 6, Acuity: 3.0
Demographics: M, WHITE, Transport: WALK IN
Diagnosis: PRESSURE ULCER OF SACRAL REGION, UNSPECIFIED STAGE (L89159)
Medication: VANCOMYCIN | ETC: ANALGESIC OR ANTIPYRETIC NON-OPIOID
Additional Med (Pyxis): OXYCODONE (IMMEDIATE REL 5MG TAB @ 2187-05-11 15:00:00+00:00
Vitals Time: 2187-05-12 19:25:00+00:00, HR: 78.0
Chief Complaint: ABSCESS, TRANSFER
Vitals → Temp: 97.8°F, HR: 79.0 bpm, RR: 16.0, BP: 108.0/60.0, O2: 96.0%
Pain: 3, Acuity: 3.0
Demographics: M, WHITE, Transport: AMBULANCE
Diagnosis: PRESSURE ULCER OF SACRAL REGION, UNSPECIFIED STAGE (L89159)
Medication: *MULTIVITAMIN WITH IRON AND FOLIC ACID | ETC: NAN
Additional Med (Pyxis): VANCOMYCI 1000MG/200ML 200ML BAG @ 2139-08-07 06:14:00+00:00
Vitals Time: 2139-08-07 05:23:00+00:00, HR: 92.0

Question: What is reasoning behind prescribing vancomyci to a patient?

Answer: The patient is a candidate for vancomycin due to bacterial infections and the presence of a pressure ulcer and abscess.
```

- PROPOSED MODEL AND FRAMEWORK-

AGENT 2 - EMERGENCY DEPARTMENT STAY SUMMARIZER AGENT

Key points:

- Creates concise, chronological, and clinically focused summaries of patient experiences in the Emergency Department (ED).
- Enables clinicians to quickly understand patient histories, aiding timely and informed clinical decisions.

Framework Overview:

- Data Integration & Summarization:
  - Combines patient data (diagnoses, treatments, medications, vital signs) into structured narratives.
- Fine-Tuning and Transfer Learning:
  - Uses knowledge gained from fine-tuning NLP models on medical data (NCBI Disease Corpus) to inform summarization methods.

MODEL & FOUNDATIONAL NLP DETAILS AND RATIONALE

- BioClinicalBERT Supervised NER:
  - Fine-tuned on NCBI Disease Corpus, achieving high precision (0.77) and recall (0.84). Provided baseline performance and methodological insights for effective clinical entity extraction in summarization.
- Mistral-7B Prompt Engineering (CoT):
  - Evaluated zero-shot, few-shot, and Chain-of-Thought prompts on NCBI dataset. Gained insights into effective prompt design for clinical summarization and entity extraction tasks.
- PEFT (LoRA) Fine-Tuning Exploration:
  - Identified critical resource and technical requirements for efficient fine-tuning of summarization models, informing hardware and software decisions for implementing PEFT methods.

MODEL EVALUATION

◆ **Automated Metrics**

- ROUGE (1/2/L), BLEU, METEOR – Text similarity to expert-written summaries
- BERTScore – Semantic similarity via embeddings
- Loss curves – Convergence & overfitting tracking

◆ **Clinical Accuracy**

- Entity-level precision/recall: Diagnoses, meds, vitals
- Temporal accuracy: Sequence of events preserved
- Factual consistency: No hallucination, structured checks

◆ **Retrieval Performance (RAG)**

- MRR, Precision@k – Quality of retrieved patient records
- Latency & throughput – Suitability for real-time ED use

FUTURE WORK

1. Expand Agent Capabilities

(a) Clinical Explanation Agent

- Faster Retrieval and Response
- Enhanced Accuracy and Relevance

(b) ED Summarizer Agent

- Implement finetuned model on ED dataset
- Get accurate output for patient summary with proper summary

(c) ED stay follow-up Agent

- Design algorithms to generate personalized discharge instructions tailored to patient demographics, diagnoses, and clinical history.
- Implement natural language generation to clearly communicate individualized care plans to patients.

2**. Generalizability & Fairness Audits**

Thank you