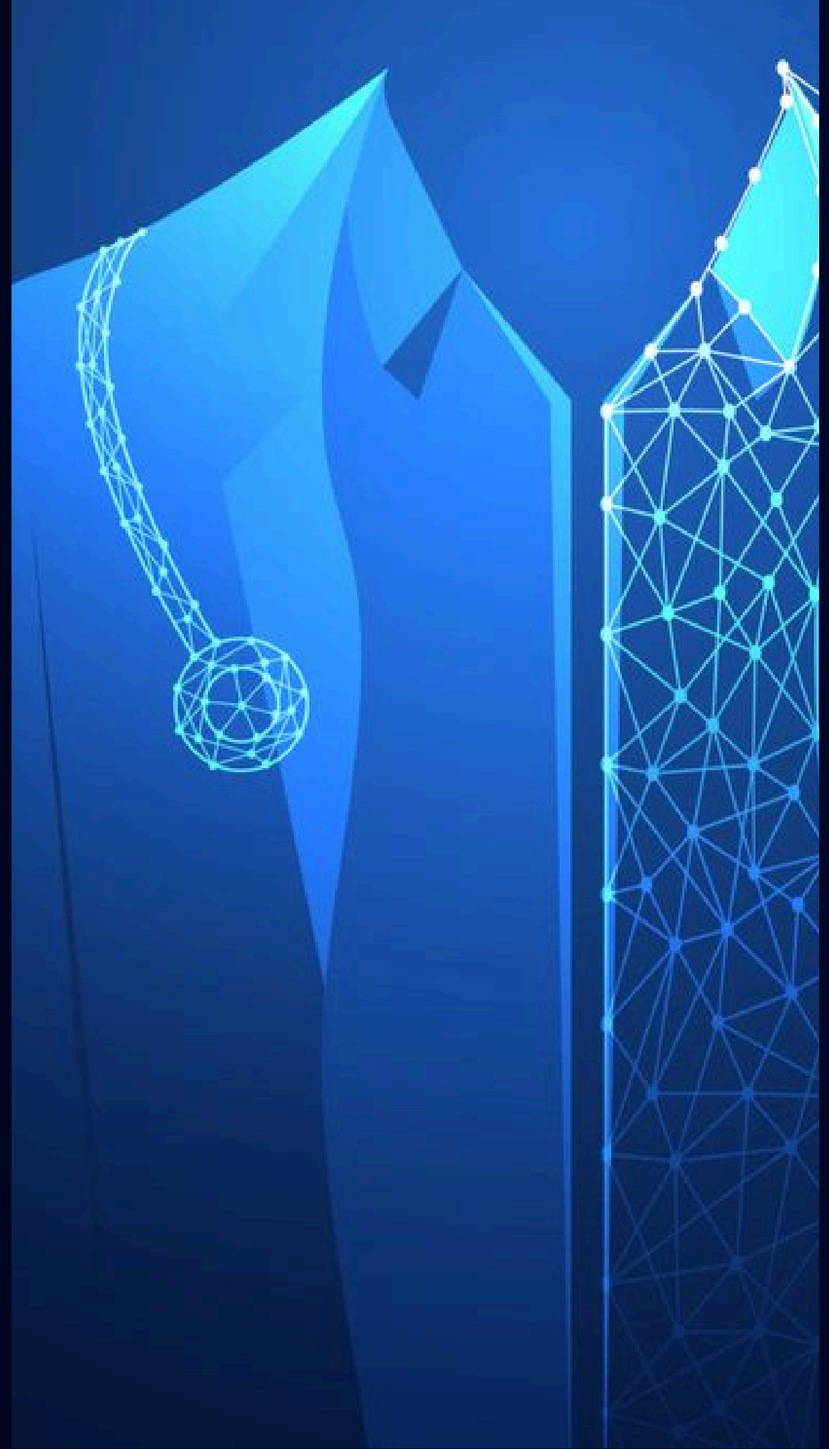


Prepared by: Ilana Estrin & Tomer Atia

MEDSUM MODEL

LM-powered summarization of medical records

USE CASE

BACKGROUND

Patient information is scattered across clinic visits, messages, lab/imaging reports, ER notes, and both digital and handwritten documents in varying formats.

WHY THIS PROBLEM MATTERS?

Heavy fragmentation makes it hard for clinicians to quickly understand the patient's course. When doctors are tired, rushed, or unfocused, the risk of missing critical information, confusion, or clinical errors increases.

WHY IT'S CHALLENGING ?

Clinical texts come from multiple sources, with different lengths, styles, levels of structure, and inconsistent terminology - some digital, some scanned, some handwritten.

HOW IT'S SOLVED TODAY?

Clinicians manually read through many documents, digital and written, in a slow, overloaded process that does not guarantee identification of all important events.

PROJECT TASK DESCRIPTION

PROBLEM STATEMENT

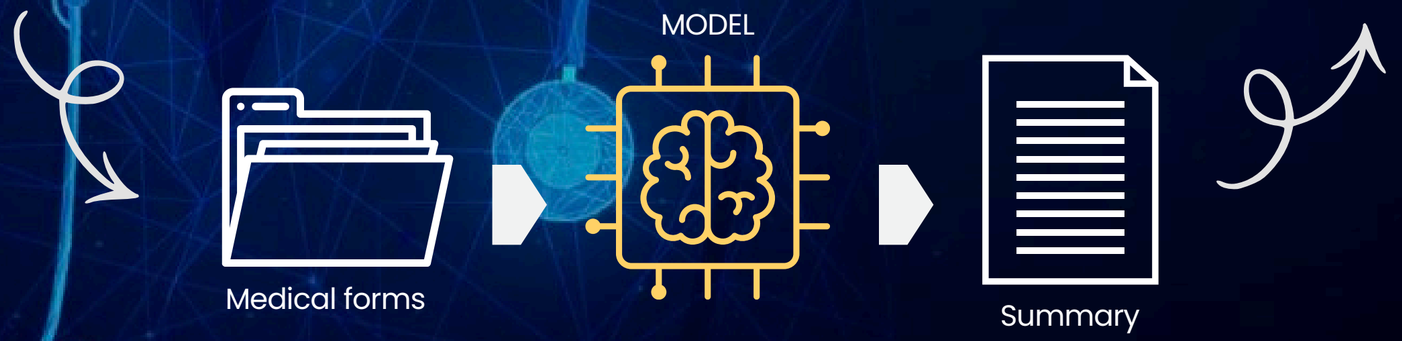
Summarizing a patient's longitudinal medical history from multiple clinical documents into one coherent, structured "patient journey" narrative.

INPUT

Ordered set of patient-related clinical texts {visit notes, doctor–patient messages, ER notes, lab/imaging reports}, all chronologically aligned on a single timeline.

OUTPUT

A concise 3–7 sentence longitudinal **"patient journey" summary** covering diagnoses, treatments, key clinical events, timeline progression, and current health status.



NOVELTY

Multi-document longitudinal summarization
Extraction → Structuring → LLM Summary
Handling sensitive, inconsistent, multi-source clinical text formats

MODELS & METHODS

MODELS

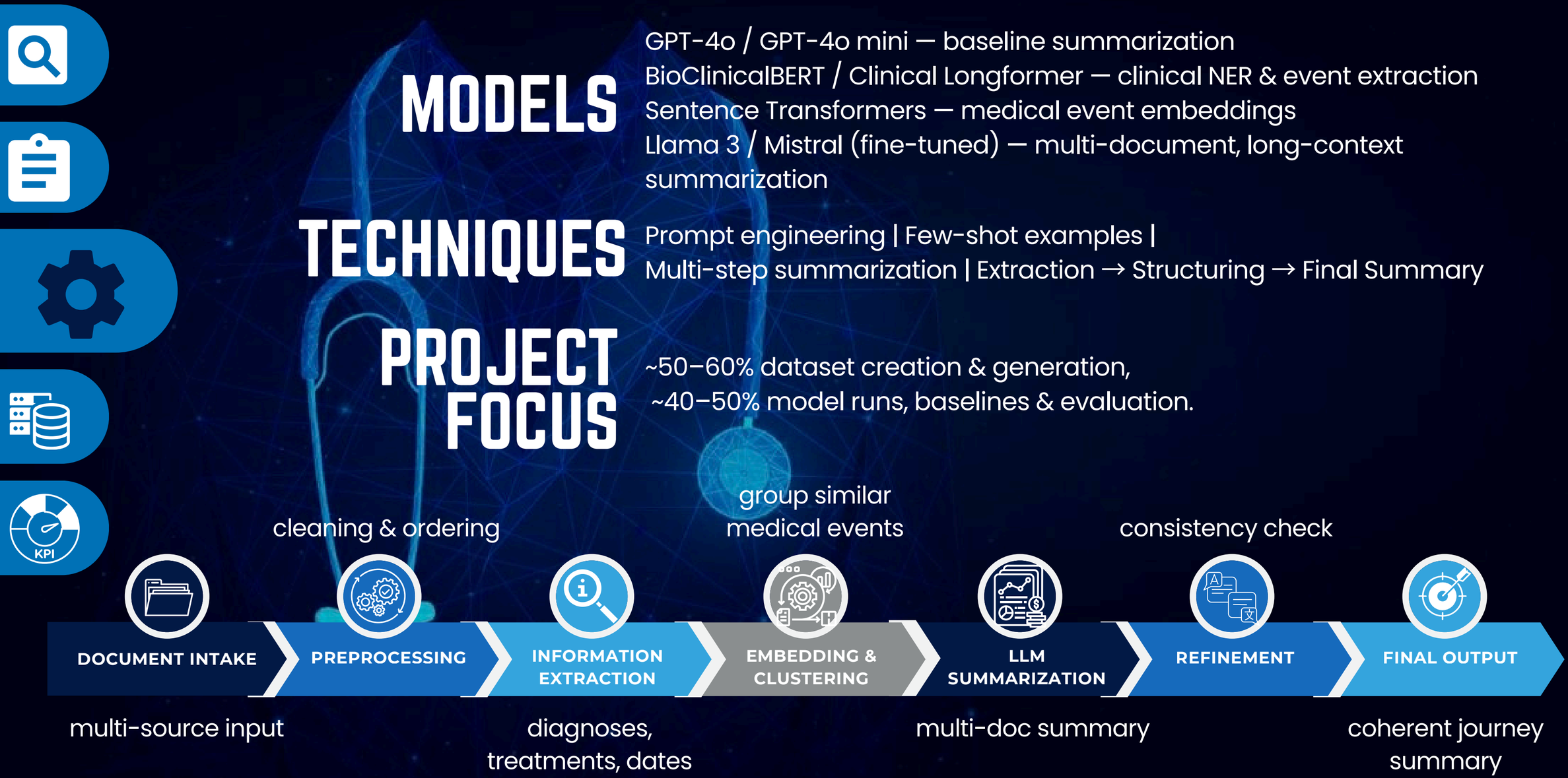
GPT-4o / GPT-4o mini — baseline summarization
BioClinicalBERT / Clinical Longformer — clinical NER & event extraction
Sentence Transformers — medical event embeddings
Llama 3 / Mistral (fine-tuned) — multi-document, long-context summarization

TECHNIQUES

Prompt engineering | Few-shot examples |
Multi-step summarization | Extraction → Structuring → Final Summary

PROJECT FOCUS

~50–60% dataset creation & generation,
~40–50% model runs, baselines & evaluation.



DATA SPECIFICATION & GENERATION

DATA REQUIREMENTS

Longitudinal clinical documents aligned on a single timeline,
including visit notes, doctor–patient messages, ER notes,
and lab/imaging reports in mixed formats (free text + structured).

DATA SOURCES

MIMIC-III | MIMIC-IV | i2b2
Synthetic patient timelines generated using GPT-4o.

DATA GENERATION

Creation of synthetic multi-visit patient histories, including diagnoses,
treatments, medications, dates, and lab/imaging events.
Clinical event extraction used to produce structured labels.

no full manual labeling → self-supervised + synthetic labels

METRICS & KPIS

3–7 sentences
Summary Length

Minimum Performance Targets

80% ⩽ Key event coverage	90% ⩽ Chronological Order	0.35 ⩽ ROUGE-L
0.85 ⩽ BERTScore	95% ⩽ Factual Accuracy	4/5 ⩽ Human Readability Score

Measurement Protocol: Comparison to structured synthetic ground truth using ROUGE-L, BERTScore, coverage, chronology, and factual accuracy.
Ground Truth: Structured summaries created during synthetic data generation.