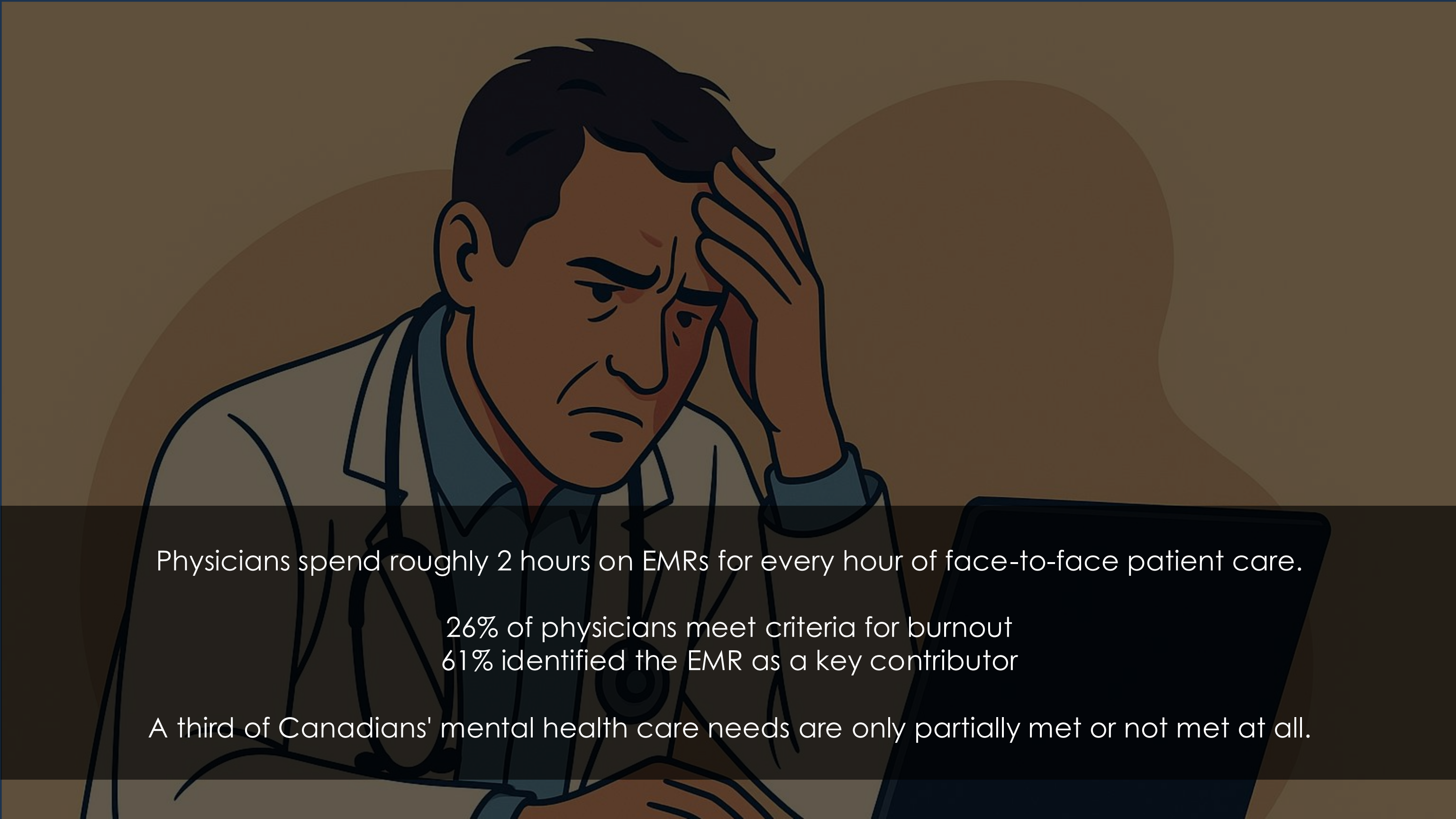


Human-centered AI for Mental Health: From Risk Prediction to Physician-aligned Design Guidelines

Syed Muhammad Ibne Zulfiker

**Dalla
Lana**
School of
Public Health



An illustration of a male physician with dark hair, wearing a white lab coat over a blue shirt. He has a stethoscope around his neck and is sitting at a desk with a laptop. He is looking down at the laptop with a distressed expression, his right hand pressed against his forehead. The background is a solid brown color with a large, faint, circular shape behind him.

Physicians spend roughly 2 hours on EMRs for every hour of face-to-face patient care.

26% of physicians meet criteria for burnout
61% identified the EMR as a key contributor

A third of Canadians' mental health care needs are only partially met or not met at all.

Thesis Statement

It is possible to more effectively adapt **modern AI techniques** for **mental health applications** that support clinicians by:

- (1) **quantitatively enhancing early assessment performance through the integration of LLM embeddings with structured features, and**
- (2) **qualitatively identifying physician' needs and capturing their cognitive and diagnostic processes to ensure that AI-driven summaries are presented in a manner consistent with their practice.**

Thesis Statement

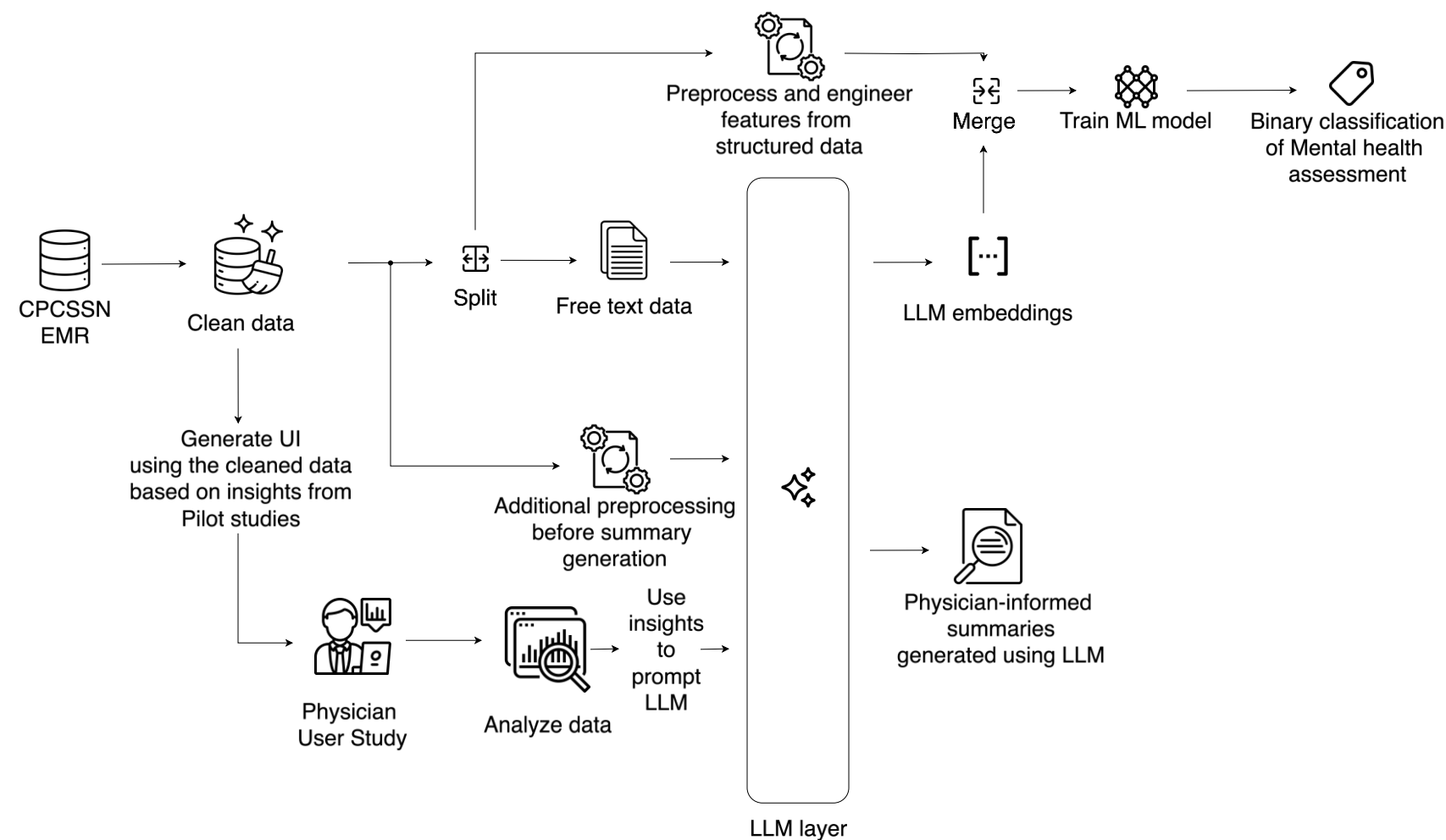
It is possible to more effectively adapt **modern AI techniques** for **mental health applications** that support clinicians by:

(1) Study 1: **Augmenting Early Mental Health Assessment Using LLM Embeddings**

(2) Study 2: **Formulating Physician-aligned Design Guidelines for AI Summarization**

System Overview

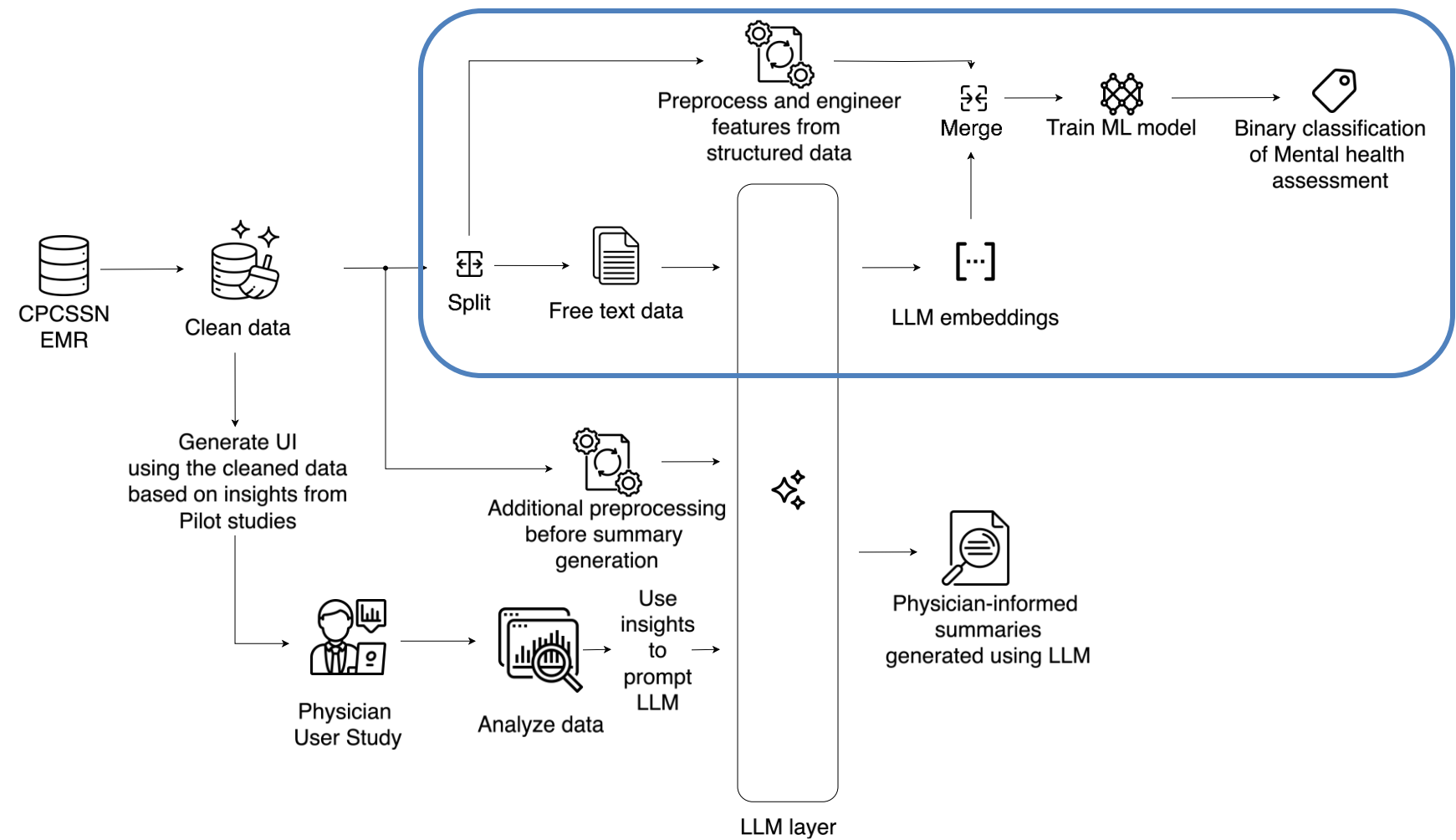
(1) Study 1: Augmenting Early Mental Health Assessment Using LLM Embeddings



(2) Study 2: Formulating Physician-aligned Design Guidelines for AI Summarization

System Overview

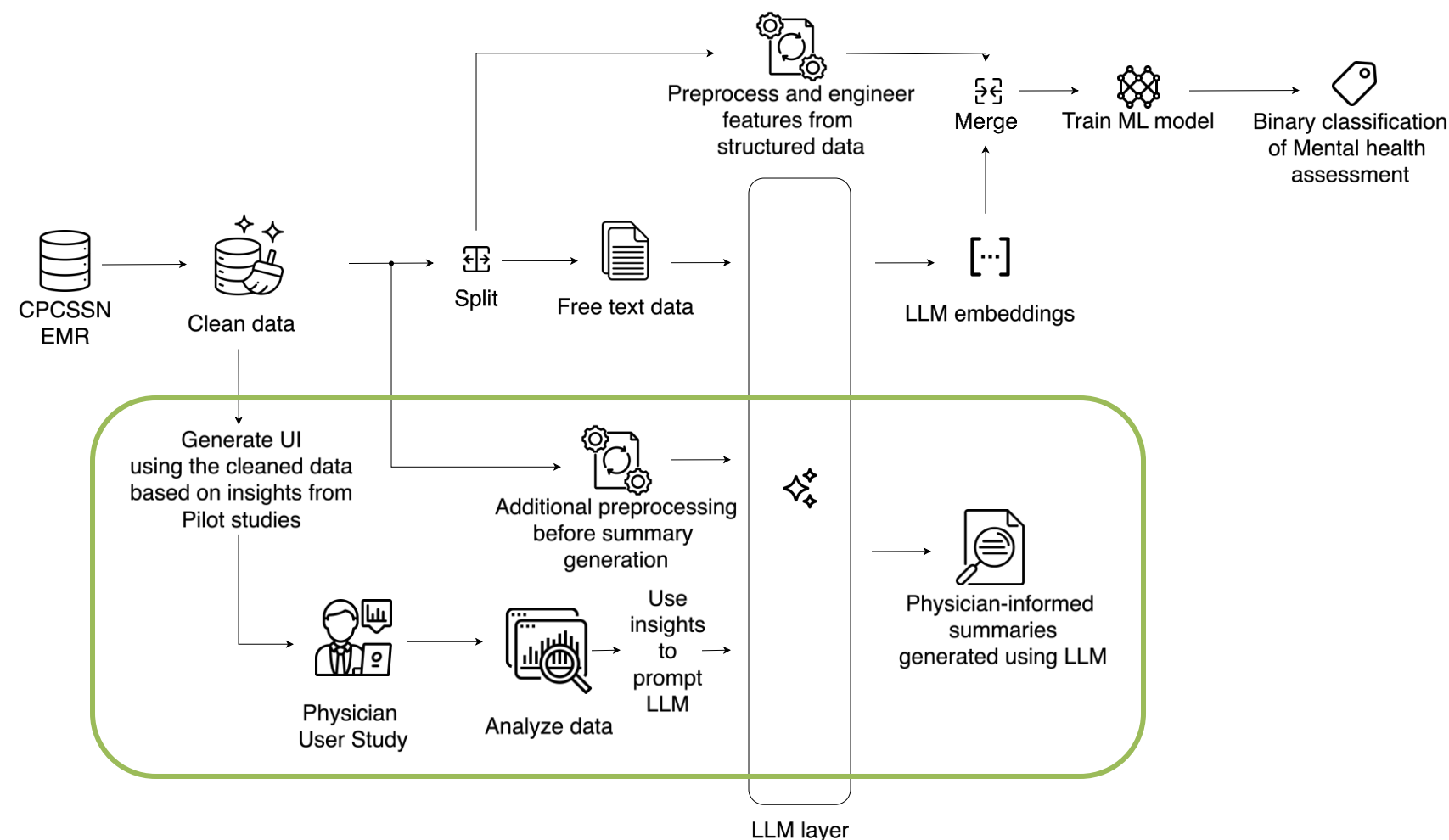
(1) Study 1: Augmenting Early Mental Health Assessment Using LLM Embeddings



(2) Study 2: Formulating Physician-aligned Design Guidelines for AI Summarization

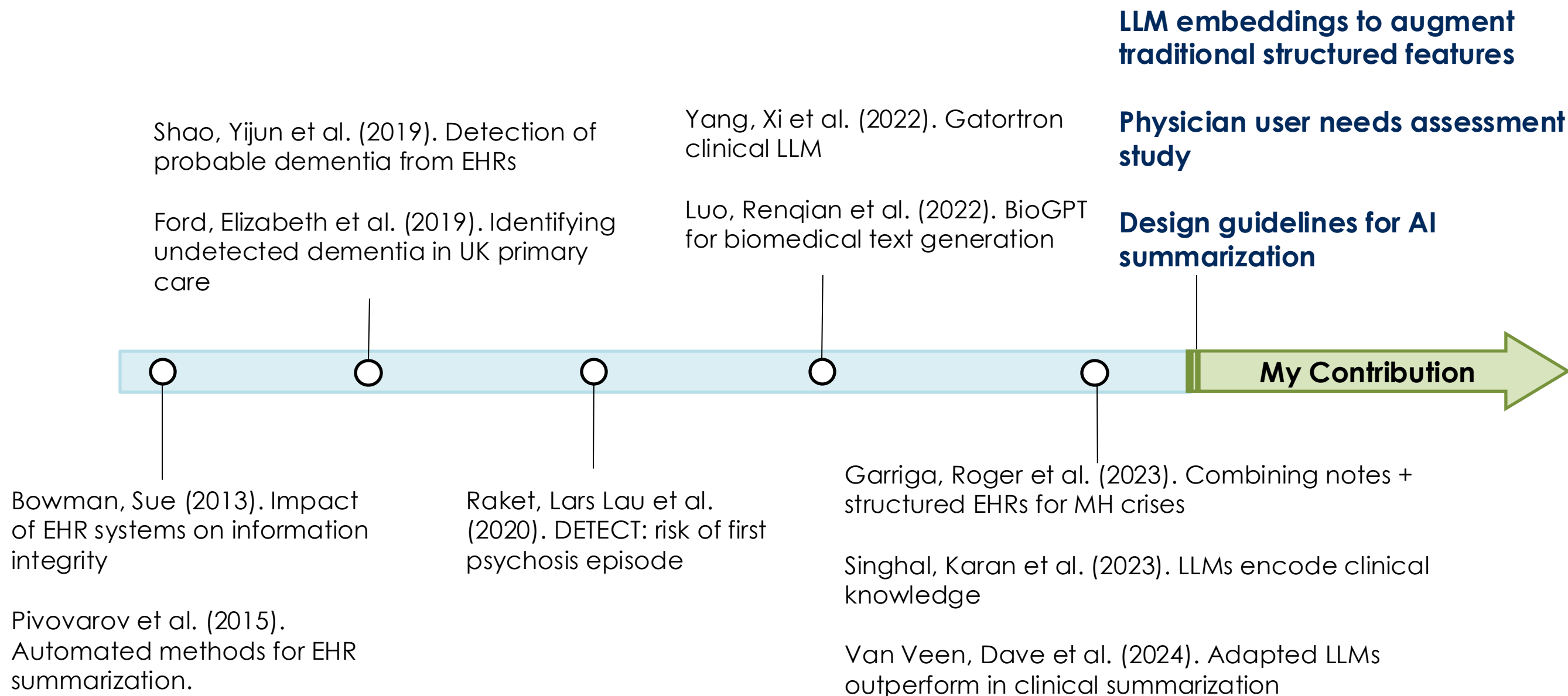
System Overview

(1) Study 1: Augmenting Early Mental Health Assessment Using LLM Embeddings



(2) Study 2: Formulating Physician-aligned Design Guidelines for AI Summarization

Background/Prior Work – Alternative Example



Study 1: Augmenting Early Mental Health Assessment Using LLM Embeddings

- **RQ1 - Modality Contribution:** What is the incremental benefit of augmenting a structured-data-only classifier with LLM-derived text embeddings?
- **RQ2 - Embedding Dimensionality:** How does classification performance change as we vary the size of the text embeddings?
- **RQ3 - Prediction Lead Times:** How does model performance evolve when enforcing different lead times?

CPCSSN Dataset

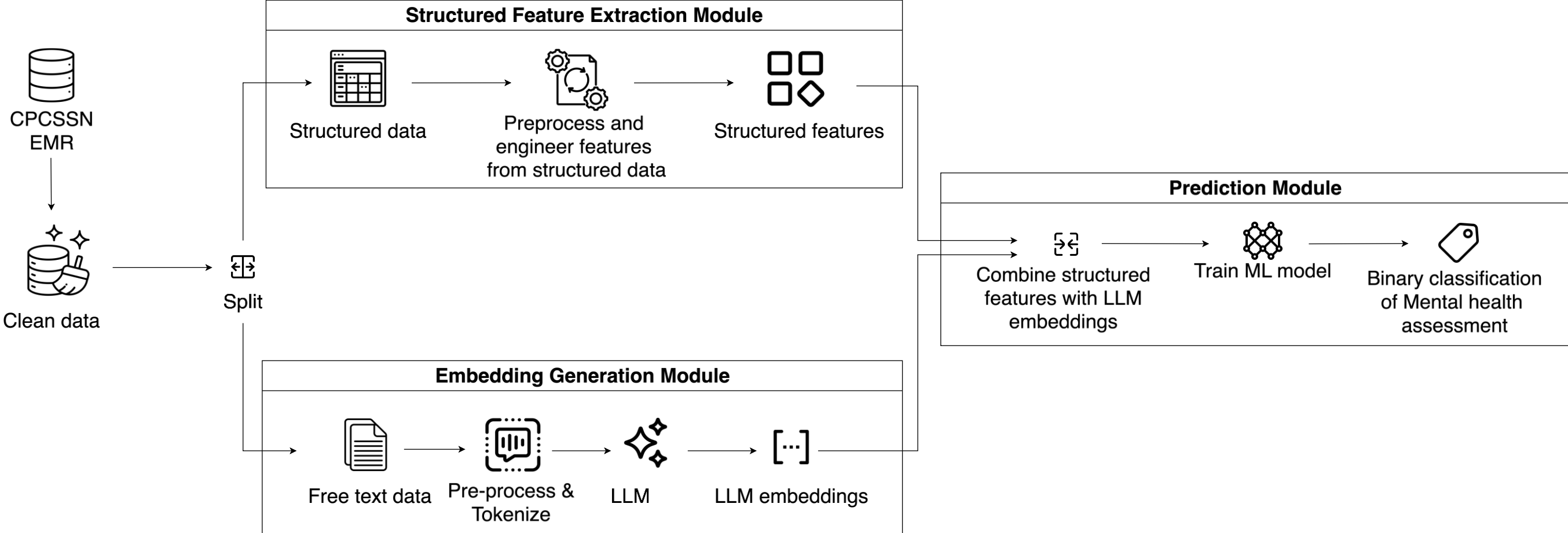
A pan-Canadian EMR database comprising routine primary care information including:

1. Demographic Information
2. Clinical Observations and Examination Results:
3. Laboratory Test Results (Biomarkers)
4. Medication Information
5. Encounter Information
6. Allergies and Risk Factors
7. Medical Procedure Information

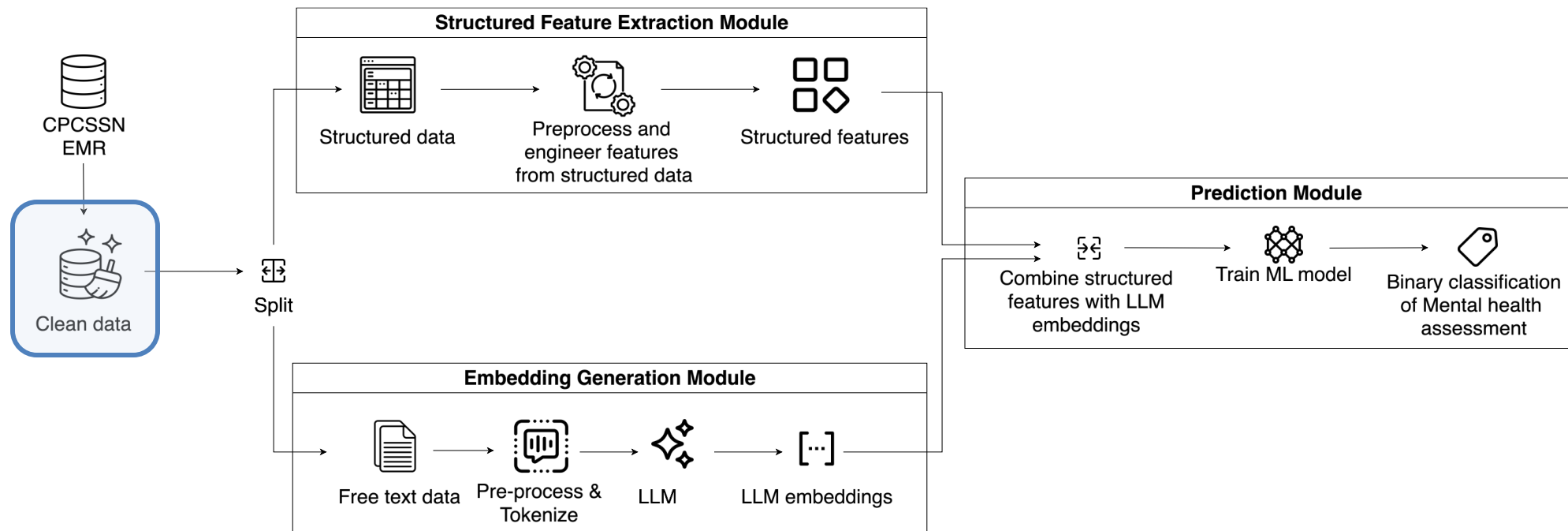
CPCSSN Dataset

- **352,161** total distinct patients
- **40.8% Male, 59.2% Female**
- **12** distinct healthcare networks (1 held out as **unseen test set**)
- Training set had the following class distribution:
 - **19.16% patients with MHD**
 - **80.84% without any MHD**

System Architecture

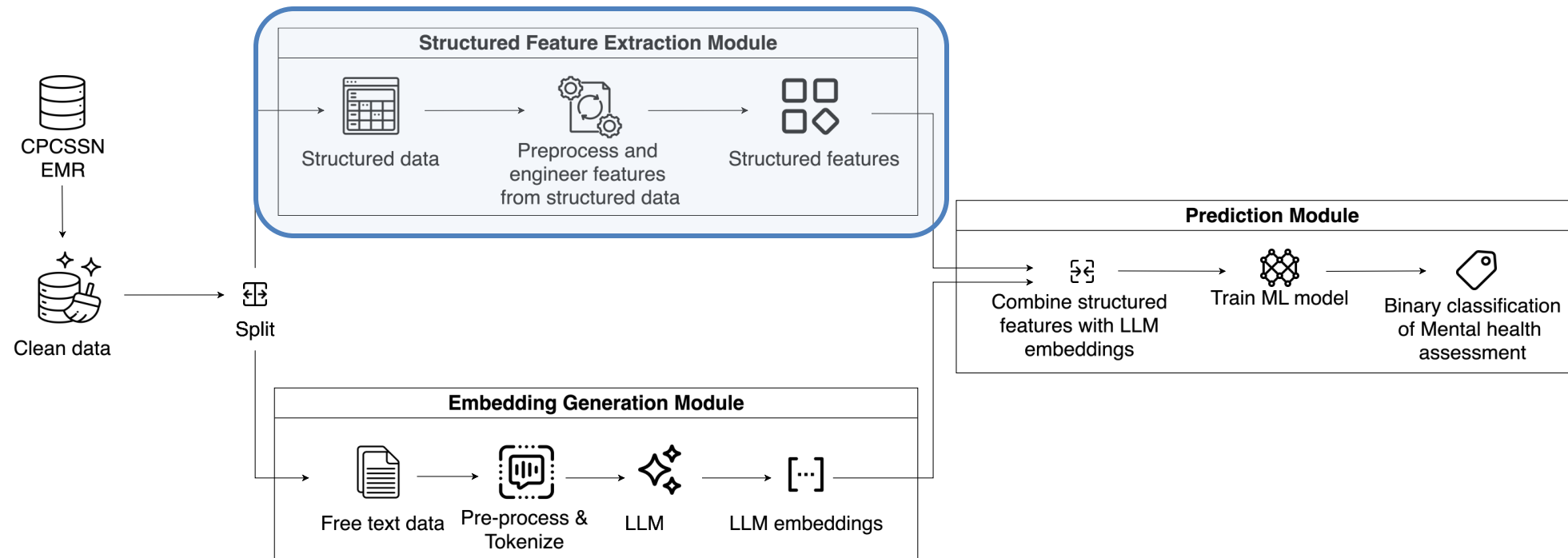


System Architecture



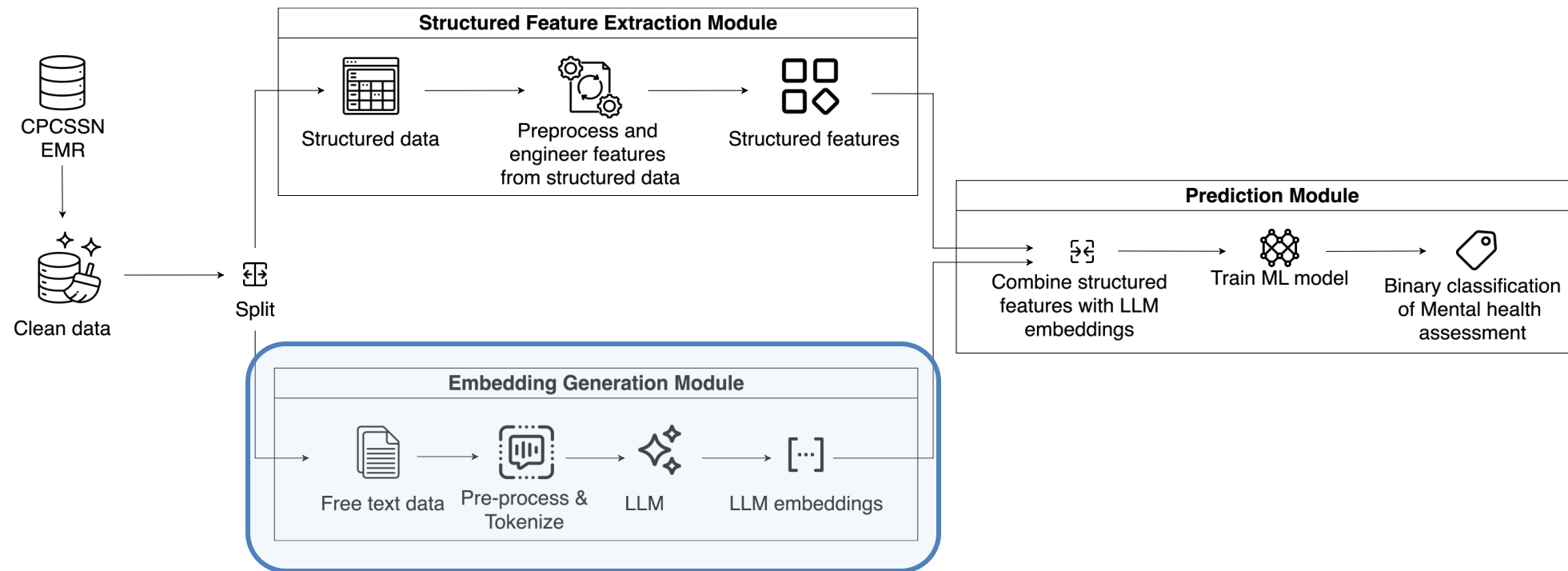
- **Outlier** Removal
- Removing **Duplicates**
- Generating **Calculated Features**
- Handling **Missing Values**
- **Validating** Data Across Tables
- Merging All Tables into a **Single Master Dataframe**
- **Standardizing** Fields

System Architecture



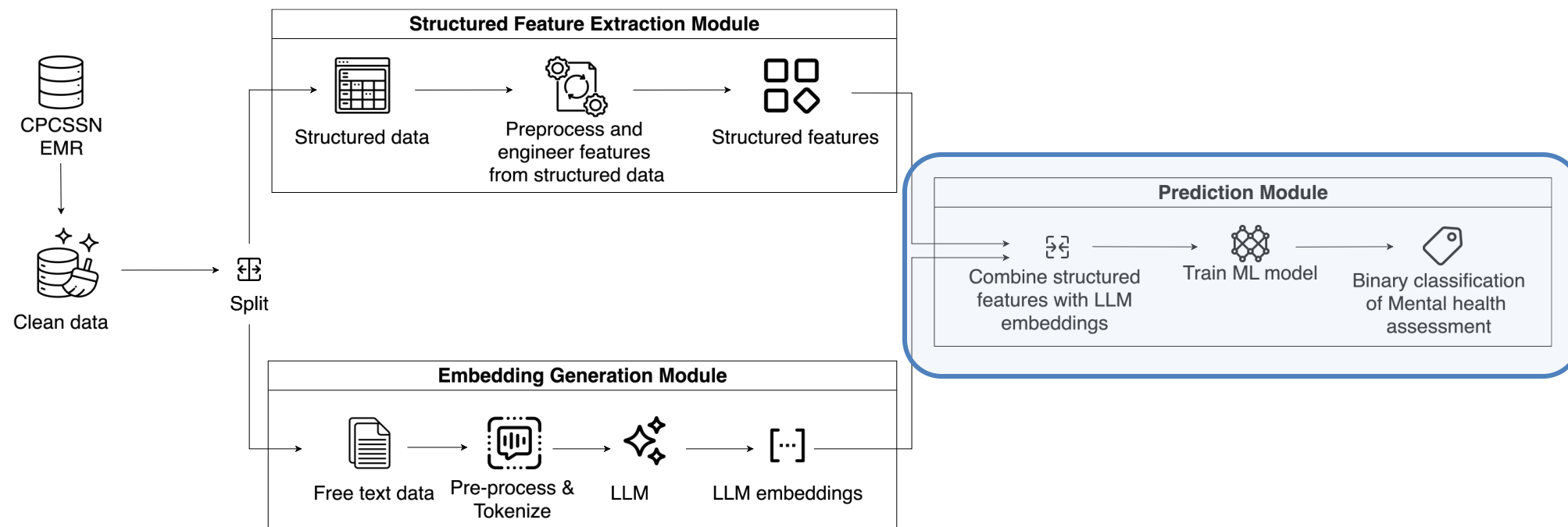
- Informed by established epidemiology and psychiatric research
- **28 structured features** including variables encoding:
 - **Demographics**
 - **Health service engagement**
 - **Comorbidities**
 - **Medications**
 - **Risk factors**
- **Multicollinearity** (Pearson correlations > 0.7)
- **MICE imputation**

System Architecture



- Open-source **Qwen 2.5 7B**
- Clinical Notes **concatenated**
- **Tokenized**
- **Truncated** for lead times
- **Embeddings** generated using a cloud GPU instance
- **3584-dimensional vectors**
- Reduced using **PCA**

System Architecture



- KNN, LR, ADA, RF, XGB, LGBM
- **Hyperparameter tuning** using 5-fold cross validation on AUC
- **Random Undersampling**
- **10-fold stratified cross-validation** for comparative analysis
- **ROC-AUC, recall, accuracy, precision, F1**
- **Pairwise t-test** with Holm-Bonferroni correction

Results

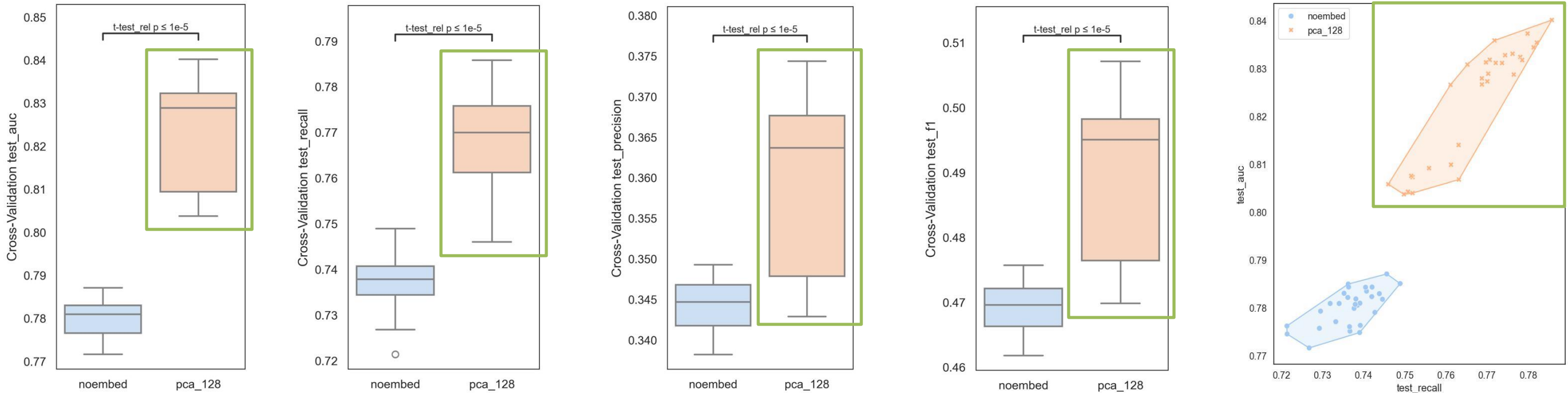
Architecture Comparison

Model (128-dims)	Accuracy (%)	AUC (%)	Recall (%)	Precision (%)	F1 (%)
LR [baseline]	68.76	70.25	58.58	29.26	39.03
KNN	82.17***	68.37***	17.53***	44.46***	25.14***
ADA	69.81***	78.13***	72.09***	32.61***	44.91***
RF	71.42***	80.73***	75.45***	34.55***	47.40***
XGB	73.16***	82.99***	77.16***	36.46***	49.52***
LGBM	73.47***	83.37***	77.59***	36.84***	49.95***

Results

RQ1 - Modality Contribution

Model	Accuracy (%)	AUC (%)	Recall (%)	Precision (%)	F1 (%)
LGBM noembed [baseline]	68.22	77.95	73.17	34.49	46.88
LGBM 128- dimensional embedding	73.47***	83.37***	77.59***	36.84***	49.95***



Results

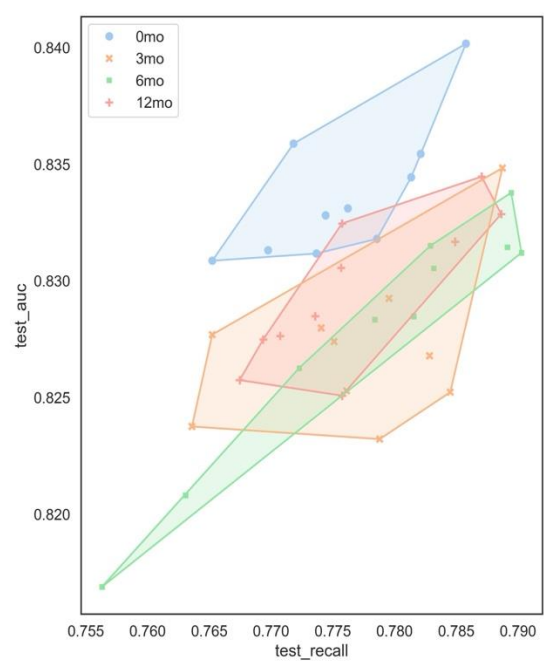
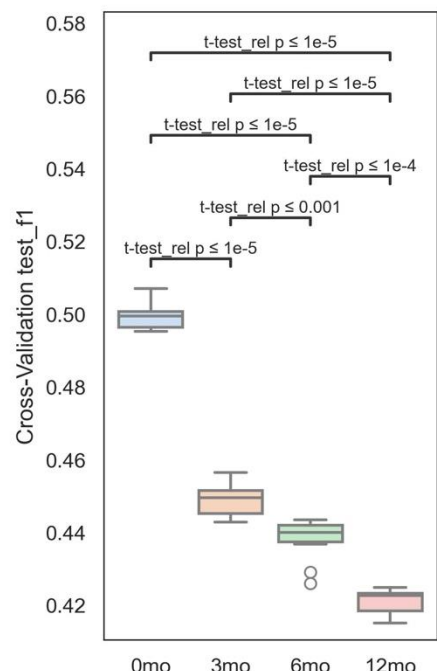
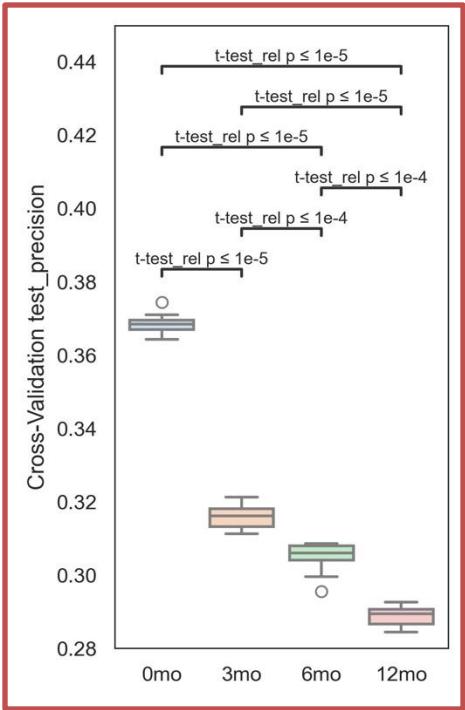
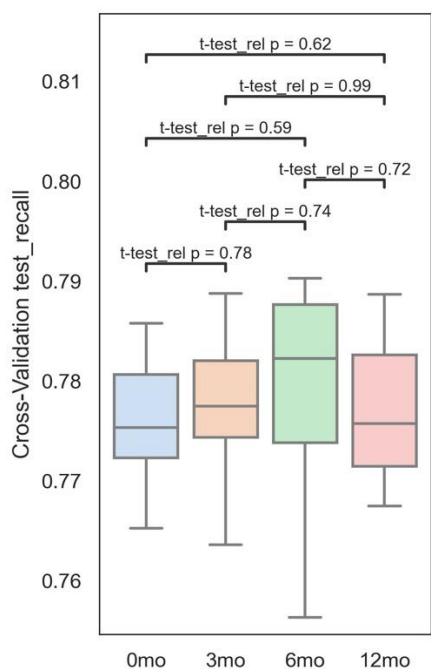
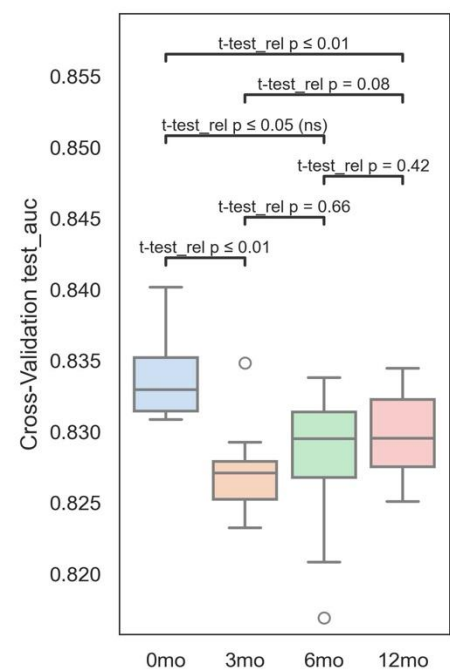
RQ2 - Embedding Dimensionality

Model (dimensions)	Accuracy (%)	AUC (%)	Recall (%)	Precision (%)	F1 (%)
LGBM (128) [baseline]	73.47	83.37	77.59	36.84	49.95
LGBM (256)	73.31**	83.24**	77.82	36.71*	49.88
LGBM (512)	73.11***	83.04***	77.41	36.45***	49.56***

Results

RQ3 - Prediction Lead Times

Model (lead time in months)	Accuracy (%)	AUC (%)	Recall (%)	Precision (%)	F1 (%)
LGBM (0) [baseline]	73.47	83.37	77.59	36.84	49.95
LGBM (3)	72.40***	82.71**	77.69	31.58***	44.90***
LGBM (6)	72.39***	82.79*	77.87	30.48***	43.81***
LGBM (12)	72.55***	82.96**	77.69	28.89***	42.12***



Results

Unseen test set

Dataset	Accuracy (%)	AUC (%)	Recall (%)	Precision (%)	F1 (%)
Train set 10-fold CV	73.47	83.37	77.59	36.84	49.95
Holdout test set	73.51	83.23	77.65	36.88	50.01

Study 2: Formulating Physician-aligned Design Guidelines for AI Summarization

We set out to elicit physician perspectives across 3 domain:

- **Data Priority:** The prioritization of clinical data within EMRs to understand patient profiles during consultation
- **Summary Structure:** The preferred sequence and format for AI-generated patient summaries
- **Value Proposition of AI:** The perceived value and utility of AI-generated summaries in improving clinical workflows

23

Study 2: Formulating Physician-aligned Design Guidelines for AI Summarization

- Feasibility Pilot Study
- Designing a Scenario and a physician EMR interface
- Revising Study Protocol
- User Needs Assessment Study

Feasibility Pilot Study

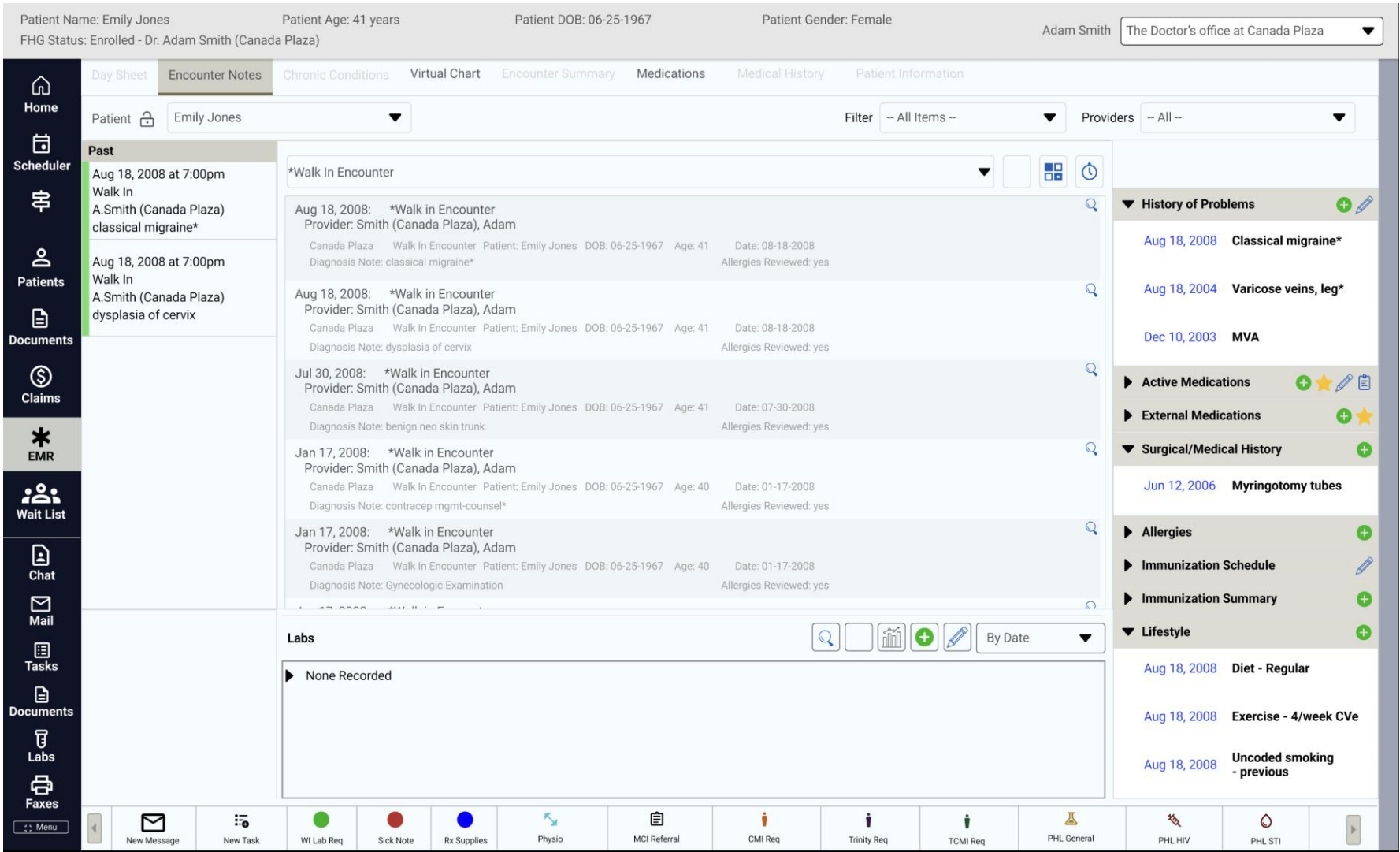
- A licensed physician with **30 years of clinical experience**
 - Scrollable PDFs
 - **Review the EMRs** and synthesize into clinically meaningful **summaries**
 - **Think aloud** and voice any concerns regarding the study protocol
-
- Importance of **Presenting Complaint (Scenario)**
 - Need for an **interactive interface (along with a reference production EMR)**
 - **Information Seeking Behavior** and **Ideal Summary Structure**
 - **Information fatigue** synthesizing **multiple patient records**

Designing a Scenario

- New patient **presenting with anxiety issues**
- Limiting synthesis to a **single patient record** to **mitigate cognitive overload**.
- A patient from CPCSSN who developed **anxiety** at a later stage of their medical history
- Patient has been **flagged** to be at **risk of Mental Health Disorders** by our **early screening model**.
- Physicians presented with medical records **up to, but not including, the first recorded mental health diagnosis**.

Designing a Physician EMR UI

Closely modeled after the reference production EMR using Figma



Study Protocol

2 phases

- **Phase 1:**

Semi Structured sessions (n=3) involving:

- Demographic & Background Survey
- EMR Review with Think-Aloud
- Post-Task Questionnaire

- **Phase 2:**

Outside the scope of this thesis but included to inform future direction:

- Blind Quality Assessment using an adapted PDQI-10 scale
- Usefulness & Acceptance Survey using UTAUT validated scale

Analysis Approach

- **Transcribe** interview transcripts
- Develop **Codebook**
- **2 independent coders** apply codebook to transcripts
- A **third independent coder** conducts consensus coding for codes with $\kappa < 0.8$
- Develop **themes** from the codes using a semi-deductive approach
- **3 themes** emerged:
 1. **Information Seeking Behavior**
 2. **Summary Structure**
 3. **Value Proposition of AI**

Theme 1: Information Seeking Behavior

- **Prioritization of Demographics and Past Medical History**
- **Filtering By Presenting Complaint**
- **Lab Results and Allergies are Context Dependent**

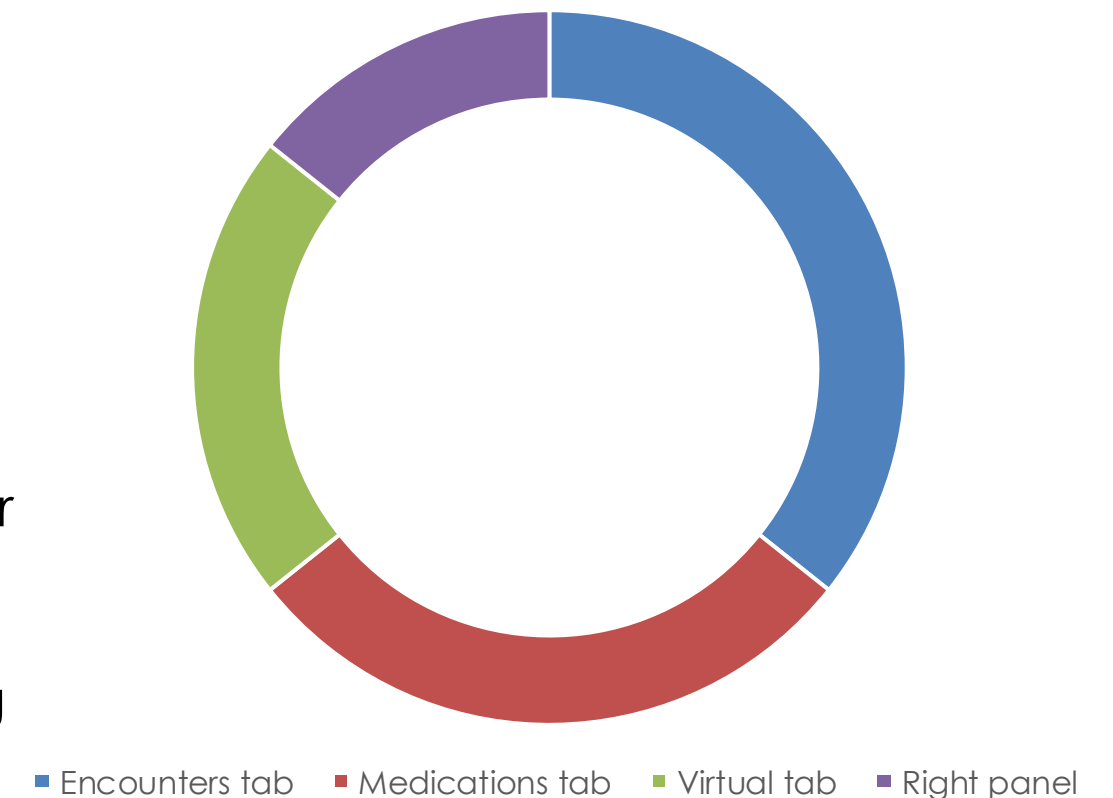
Theme 1: Information Seeking Behavior

- **Prioritization of Demographics and Past Medical History**
 - Physicians anchor clinical orientation by immediately referring to the demographics section at the top
 - Followed by patient's medical case history (chronic illnesses, major diagnoses and surgical and medical hospitalizations)
 - $\kappa = 1$ for "seeks demographics" and "seeks past medical history"
- **Filtering By Presenting Complaint**
- **Lab Results and Allergies are Context Dependent**

Theme 1: Information Seeking Behavior

- **Prioritization of Demographics and Past Medical History**
- **Filtering By Presenting Complaint**
 - High frequency of code "filters for presenting complaint"
 - Interaction patterns show filtering done on 2 key sources:
 - Encounter notes & medications (>50% of interactions)
 - At least a third of review time dedication to Encounter tab alone
 - Filtering might lead to physician inadvertently overlooking cues not explicitly linked to presenting complaint
- **Lab Results and Allergies are Context Dependent**

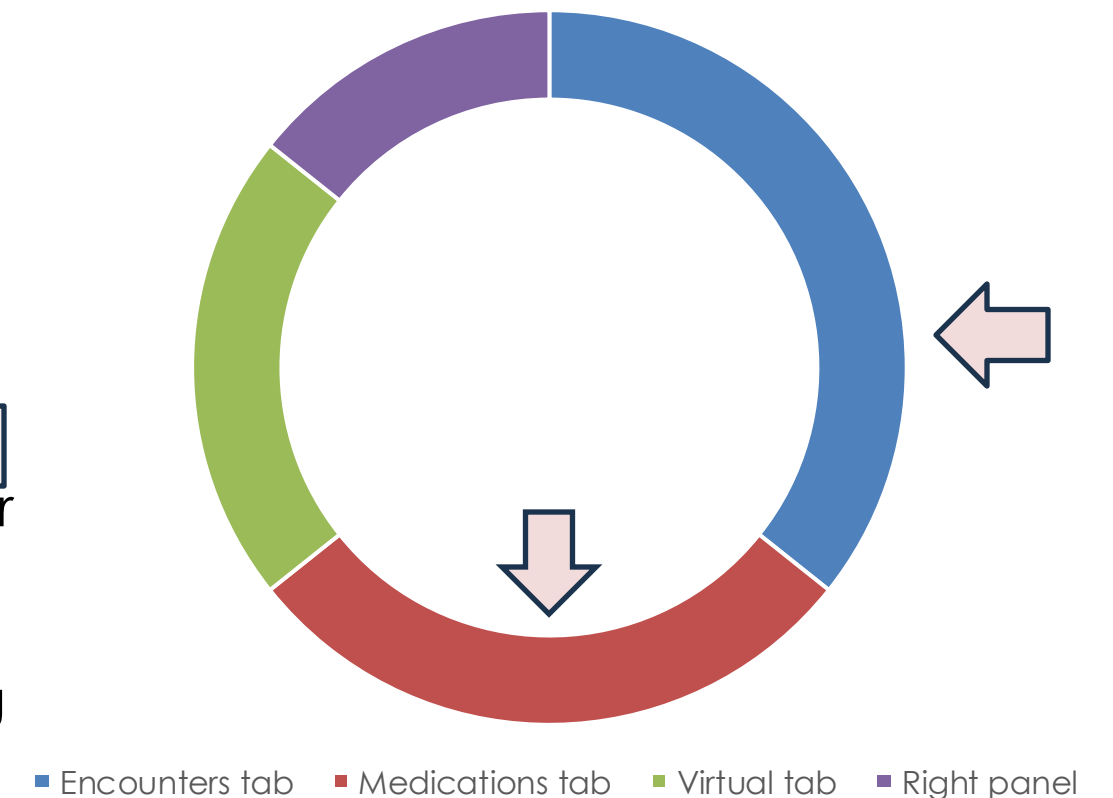
Relative Interaction Frequency



Theme 1: Information Seeking Behavior

- **Prioritization of Demographics and Past Medical History**
- **Filtering By Presenting Complaint**
 - High frequency of code "filters for presenting complaint"
 - Interaction patterns show filtering done on 2 key sources:
 - Encounter notes & medications (>50% of interactions)
 - At least a third of review time dedication to Encounter tab alone
 - Filtering might lead to physician inadvertently overlooking cues not explicitly linked to presenting complaint
- **Lab Results and Allergies are Context Dependent**

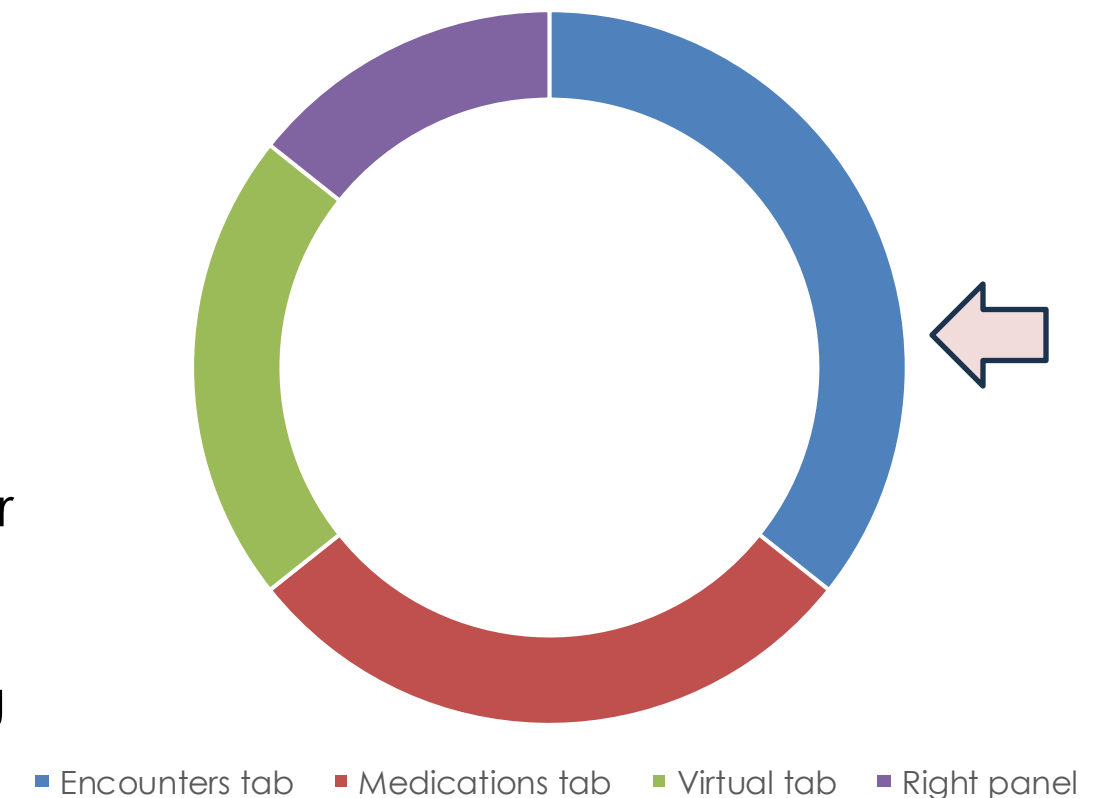
Relative Interaction Frequency



Theme 1: Information Seeking Behavior

- **Prioritization of Demographics and Past Medical History**
- **Filtering By Presenting Complaint**
 - High frequency of code "filters for presenting complaint"
 - Interaction patterns show filtering done on 2 key sources:
 - Encounter notes & medications (>50% of interactions)
 - At least a third of review time dedication to Encounter tab alone
 - Filtering might lead to physician inadvertently overlooking cues not explicitly linked to presenting complaint
- **Lab Results and Allergies are Context Dependent**

Relative Interaction Frequency



Theme 1: Information Seeking Behavior

- **Prioritization of Demographics and Past Medical History**
- **Filtering By Presenting Complaint**
- **Lab Results and Allergies are Context Dependent**
 - Only 1 of 3 physicians interacted with the Allergy section
 - Average engagement with virtual tab was below 7 secs
 - Some physicians preferred real-time review:

"I get most of my benefit doing that (reviewing lab results) with the patient while I'm in the room with them and able to share my screen..."

Theme 2: Summary Structure

- **Recency**
- **Preference of Structure Over Paragraphs**
- **Information Priority**

Theme 2: Summary Structure

- **Recency**
 - High frequency of the code "prioritizes recent info"
 - "I'm going to be looking back one or two visits....The last year is the most important..."
- **Preference of Structure Over Paragraphs**
- **Information Priority**

Theme 2: Summary Structure

- **Recency**
- **Preference of Structure Over Paragraphs**
 - "Yeah, I would like it (the summary) separated into sections"
 - "Like bullet points. So not paragraphs, but just like ... can you summarize this patient in eight lines like 8 short statements?"
- **Information Priority**

Theme 2: Summary Structure

- Recency
- Preference of Structure Over Paragraphs
- Information Priority

Participant ID	Sequence of Information Expected In a Good Summary (priority decreases left-to-right)				
F1	Demographics	Past medical history	Encounters	Medications	Lab results
P1	Demographics	Past medical history (chronic illnesses)	Medical Procedures	Medications (active medications)	Lab results (abnormal)
P2	Past medical history	Encounters	Medications	Allergies (when prescribing)	Lab results (context dependent) (specifically looks for trends)
P3	Demographics	Encounters	Medications	Lab results	Allergies (when prescribing)

Theme 2: Summary Structure

- Recency
- Preference of Structure Over Paragraphs
- Information Priority

Participant ID	Sequence of Information Expected In a Good Summary (priority decreases left-to-right)				
F1	Demographics	Past medical history	Encounters	Medications	Lab results
P1	Demographics	Past medical history (chronic illnesses)	Medical Procedures	Medications (active medications)	Lab results (abnormal)
P2	Past medical history	Encounters	Medications	Allergies (when prescribing)	Lab results (context dependent) (specifically looks for trends)
P3	Demographics	Encounters	Medications	Lab results	Allergies (when prescribing)

Theme 2: Summary Structure

- Recency
- Preference of Structure Over Paragraphs
- Information Priority

Participant ID	Sequence of Information Expected In a Good Summary (priority decreases left-to-right)				
F1	Demographics	Past medical history	Encounters	Medications	Lab results
P1	Demographics	Past medical history (chronic illnesses)	Medical Procedures	Medications (active medications)	Lab results (abnormal)
P2	Past medical history		Encounters	Medications	Lab results (context dependent) (specifically looks for trends)
P3	Demographics	Encounters	Medications	Lab results	Allergies (when prescribing)

Theme 2: Summary Structure

- Recency
- Preference of Structure Over Paragraphs
- Information Priority

Participant ID	Sequence of Information Expected In a Good Summary (priority decreases left-to-right)				
F1	Demographics	Past medical history	Encounters	Medications	Lab results
P1	Demographics	Past medical history (chronic illnesses)	Medical Procedures	Medications (active medications)	Lab results (abnormal)
P2	Past medical history	Encounters	Medications	Allergies (when prescribing)	Lab results (context dependent) (specifically looks for trends)
P3	Demographics		Medications	Lab results	Allergies (when prescribing)

Theme 2: Summary Structure

- Recency
- Preference of Structure Over Paragraphs
- Information Priority

Participant ID	Sequence of Information Expected In a Good Summary (priority decreases left-to-right)				
F1	Demographics	Past medical history	Encounters	Medications	Lab results
P1	Demographics	Past medical history (chronic illnesses)	Medical Procedures	Medications (active medications)	Lab results (abnormal)
P2	Past medical history	Encounters	Medications	Allergies (when prescribing)	Lab results (context dependent) (specifically looks for trends)
P3	Demographics	Encounters	Medications	Lab results	Allergies (when prescribing)

Theme 2: Summary Structure

- Recency
- Preference of Structure Over Paragraphs
- Information Priority

Participant ID	Sequence of Information Expected In a Good Summary (priority decreases left-to-right)				
F1	Demographics	Past medical history	Encounters	Medications	Lab results
P1	Demographics	Past medical history (chronic illnesses)	Medical Procedures	Medications (active medications)	Lab results (abnormal)
P2	Past medical history	Encounters	Medications	Allergies (when prescribing)	Lab results (context dependent) (specifically looks for trends)
P3	Demographics	Encounters	Medications	Lab results	Allergies (when prescribing)

Theme 3: Value Proposition of AI

- Interface Complexity
- Need for AI summaries

Theme 3: Value Proposition of AI

- **Interface Complexity**

- Information retrieval is tedious and time consuming:

"I find it (the EMR they routinely use) super hard and the filtering options are also complex and not easy to do ... I find the workflows (for information retrieval) in it really clunky and difficult ..."

- EMR structures can actively deter physicians from exploring potentially relevant data:

"If there was more information, if I had to scroll through multiple windows to see everything, click into notes—that really becomes a lot more time consuming, right? ... I will avoid doing that if I don't need to."

- **Need for AI summaries**

Theme 3: Value Proposition of AI

- **Interface Complexity**
- **Need for AI summaries**
 - "OK, it would be great if I didn't have to look through all this stuff and if when I open the chart, I immediately saw the summary that we just described..."
 - "I think with [an AI] summary tool, focusing on that approach that I kind of outlined and generating that kind of summary, I personally would find that to be helpful ... I mean, it might just save me a few clicks in terms of having to go through and collect all that information myself and then sort of storing it in my memory while I talk to the patient, right? ... That's the value proposition."

Study Implications

- Physicians prefer a **structured summary**
- **Demographic information** and **presenting complaint** provide anchors
- **Medical history** including chronic illnesses, major diagnoses, and surgical and medical hospitalizations are prioritized
- **Encounters and Medications** are then reviewed **filtered by presenting complaint**. **Recent** information is given priority.

Study Implications

You are an experienced physician. Given the longitudinal EMR information presented to you, generate a structured summary of the patient's records (ideally using bullet points). Include the sections outlined below. Organize the summary using the following headings, in order:

- **Demographic information** and **presenting complaint** provide anchors
- **Medical history** including chronic illnesses, major diagnoses, and surgical and medical hospitalizations are prioritized
- **Encounters and Medications** are then reviewed **filtered by presenting complaint**. **Recent** information is given priority.

Study Implications

You are an experienced physician. Given the longitudinal EMR information presented to you, generate a structured summary of the patient's records (ideally using bullet points). Include the sections outlined below. Organize the summary using the following headings, in order:

1. Demographic Information (include age, sex, and the presenting complaint)

- **Medical history** including chronic illnesses, major diagnoses, and surgical and medical hospitalizations are prioritized
- **Encounters and Medications** are then reviewed **filtered by presenting complaint**. **Recent** information is given priority.

Study Implications

You are an experienced physician. Given the longitudinal EMR information presented to you, generate a structured summary of the patient's records (ideally using bullet points). Include the sections outlined below. Organize the summary using the following headings, in order:

1. Demographic Information (include age, sex, and the presenting complaint)
 2. Past Medical History (include chronic illnesses, major diagnoses, and surgical and medical hospitalizations)
- **Encounters and Medications** are then reviewed **filtered by presenting complaint**. **Recent** information is given priority.

Study Implications

You are an experienced physician. Given the longitudinal EMR information presented to you, generate a structured summary of the patient's records (ideally using bullet points). Include the sections outlined below. Organize the summary using the following headings, in order:

1. Demographic Information (include age, sex, and the presenting complaint)
2. Past Medical History (include chronic illnesses, major diagnoses, and surgical and medical hospitalizations)
3. Most Recent Encounters and Medications (emphasize content relevant to the presenting complaint)

Summarization Outputs Pre- and Post- Physician-Informed Prompt Engineering

Manual Physician-written Summary	Conventional AI-generated Summary	Physician-informed AI-generated Summary
48 year old Female. Patient is presenting for an anxiety issue. Quick scan of chart for other visits for anxiety in the last three to four visits shows nothing related. Visits have been for migraine, cervical dysplasia and skin lesions. A quick assessment of medications shows no psychoactive drugs, and patient summary is not elucidating any red flags for anxiety.	Prompt: You are an expert medical professional. Summarize the medical records into a concise clinical note. Response: The medical records document various health conditions including classical migraine and dysplasia of the cervix, along with risk factors such as previous smoking and social alcohol use. The patient has been prescribed contraceptive medication multiple times for extended periods and has undergone gynecological examinations and procedures like myringotomy tubes.	Demographic Information Age: 48 years old Sex: Female Presenting Complaint: Anxiety Past Medical History Chronic Illnesses: · Classical migraine (diagnosed on 2008-08-18) · Dysplasia of cervix (diagnosed in multiple instances) Major Diagnoses / Surgical and Medical hospitalizations: · Varicose veins, leg (onset in 2004) · Motor vehicle accident (MVA) (in 2003) · History of high-risk sexual behavior and HPV infection Most Recent Encounters and Medications Encounters: · On 2008-08-18, the patient was diagnosed with classical migraine*. Medications: · Oral contraceptive pills like TRIPHASIL 21 TAB were prescribed on multiple occasions from 2005 to 2007. Additional Notes The patient has a detailed health record with multiple gynecological issues and past medical history. The recent encounter on 2008-08-18 highlighted the presence of classical migraine. Given her history, it is important to consider this diagnosis in relation to her current presentation of anxiety, as both conditions can have overlapping symptoms or be influenced by similar factors such as stress and hormonal changes

Summarization Outputs Pre- and Post- Physician-Informed Prompt Engineering

Manual Physician-written Summary	Conventional AI-generated Summary	Physician-informed AI-generated Summary
48 year old Female. Patient is presenting for an anxiety issue. Quick scan of chart for other visits for anxiety in the last three to four visits shows nothing related. Visits have been for migraine, cervical dysplasia and skin lesions. A quick assessment of medications shows no psychoactive drugs, and patient summary is not elucidating any red flags for anxiety.	Prompt: You are an expert medical professional. Summarize the medical records into a concise clinical note. Response: The medical records document various health conditions including classical migraine and dysplasia of the cervix, along with risk factors such as previous smoking and social alcohol use. The patient has been prescribed contraceptive medication multiple times for extended periods and has undergone gynecological examinations and procedures like myringotomy tubes.	Demographic Information Age: 48 years old Sex: Female Presenting Complaint: Anxiety Past Medical History Chronic Illnesses: · Classical migraine (diagnosed on 2008-08-18) · Dysplasia of cervix (diagnosed in multiple instances) Major Diagnoses / Surgical and Medical hospitalizations: · Varicose veins, leg (onset in 2004) · Motor vehicle accident (MVA) (in 2003) · History of high-risk sexual behavior and HPV infection Most Recent Encounters and Medications Encounters: · On 2008-08-18, the patient was diagnosed with classical migraine*. Medications: · Oral contraceptive pills like TRIPHASIL 21 TAB were prescribed on multiple occasions from 2005 to 2007. Additional Notes The patient has a detailed health record with multiple gynecological issues and past medical history. The recent encounter on 2008-08-18 highlighted the presence of classical migraine. Given her history, it is important to consider this diagnosis in relation to her current presentation of anxiety, as both conditions can have overlapping symptoms or be influenced by similar factors such as stress and hormonal changes

Limitations

- **PCA** dimensionality reduction (linear transformations)
- Exclusively the **Qwen 2.5** architecture
- **Retrospective** study (association rather than causal inferences)
- Exclusively interviewed **family physicians**
- **Single production EMR** (Accuro) as reference
- Some **static interface** components constrained physician workflows

Future Work

- Exploring **alternative dimensionality reduction techniques** (UMAP, t-SNE, or neural autoencoders)
- Exploring **additional LLM architectures**
- Expanding User Studies Across **Specialties and EMR Platforms**
- **Evaluation** of the Clinical Utility and Quality of Generated Summaries
- **Fine-Tuning** LLMs with Physician Expertise

Key Takeaways

- Demonstrated the value of **augmenting** structured EMR features **with LLM-derived embeddings** of unstructured clinical text
- Conducted a **user needs assessment study** to gain insights on **physician requirements, expectations, and underlying cognitive models** involved in the review and synthesis of EMRs
- Laid down a set of **design principles** incorporating **physician feedback**
- Established a **methodological foundation** for gathering and synthesizing physician feedback
- Taken together, we can **leverage risk signals** from the early assessment study to **surface relevant clinical information** using a physician-informed structure based on our **design guidelines**
- **Reducing information overload** and **streamlining the clinical decision-making process**

57