



Generative AI and Machine Learning in Healthcare



SONALI TAMHANKAR

Senior Clinical Data Scientist
Fred Hutchinson Cancer Center

Generative AI and Machine Learning in Healthcare



Non-profit Healthcare Ecosystem

Regulatory Chatbot



Sickle Cell Emergency Visits



Future Directions



Fred Hutchinson Cancer Center, Seattle

Fred Hutch has world class expertise in oncology as well as in infectious diseases and benign hematology (non-malignant blood disorders).



Fred Hutch and Seattle Cancer Care Alliance unite, reshape relationship with UW Medicine

Restructure promises to boost research and speed new cancer treatments from bench to bedside

APRIL 1, 2022 • BY FRED HUTCH NEWS SERVICE STAFF

Translating discoveries into cures, our pioneering research has saved hundreds of thousands of lives worldwide.

3

Nobel Prize Winners

Researchers from Fred Hutch have been honored with the Nobel Prize in physiology or medicine.

60+

Countries Where We Do Research

Our science takes place around the world and has an impact on people everywhere.

5700

Employees

Collaborating across the globe, our dedicated, talented people are driven by an ambitious vision.

16+

Acres of Campus

Our beautiful South Lake Union campus features cutting-edge research facilities that promote cross-disciplinary collaboration.

53K+

Individual Patients

Nearly 120,000 outpatient visits

945

Participants

In therapeutic clinical trials along with 385 adult blood stem cell transplants

10K

Screenings

By the mammogram van around Seattle and Puget Sound area

Enterprise Analytics at Fred Hutch

- The research use cases are supported by respective lab resources.
- On the care delivery side, analytics support for operational, strategic, and clinical initiatives is provided by the Enterprise Analytics, under Quality Improvement umbrella.
- The Enterprise Analytics team builds dashboards tracking key performance indicators and clinical outcomes using SQL and tableau.
- We use an Electronic Health Record (EHR) system by Epic. We have Epic experts with deep EHR expertise and an educator who helps clinicians access information using built-in EHR reporting tools.
- We have clinical notes in our SQL data mart, which can be mined when information is not available in structured format.
- Business partner bandwidth can be severely limited in a healthcare delivery organization.
- My goal is to use advanced analytics tools (including machine learning and Generative AI) to support business partner endeavors.

Generative AI and Machine Learning in Healthcare



Non-profit Healthcare Ecosystem

Regulatory Chatbot



Sickle Cell Emergency Visits

Future Directions





American Hospital
Association™

Healthcare Regulations Context

"The pervasive nature of health care regulation stems from the fundamental concerns that are at stake", "However, the present regulatory structure is neither uniform nor consistent."

- Robert I. Field, the author of Health Care Regulation in America: Complexity, Confrontation and Compromise

Project: Create a semantic search based tool to identify relevant references from multiple sources to support our Regulatory Compliance team.

1. Health systems, hospitals and PAC providers must comply with 629 discrete regulatory requirements across nine domains.

These include 341 hospital-related requirements and 288 PAC-related requirements. The four agencies that promulgated these requirements – the Centers for Medicare & Medicaid Services (CMS), the Office of Inspector General (OIG), the Office for Civil Rights (OCR) and the Office of the National Coordinator for Health Information Technology (ONC) - are the primary drivers of federal regulation impacting these providers. However, providers also are subject to regulation from other federal and state entities which are not accounted for in this report.

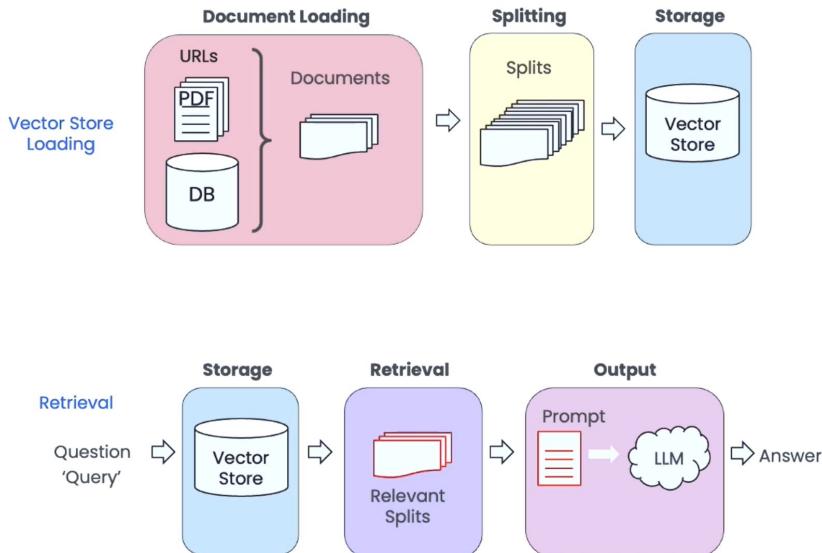
2. Health systems, hospitals and PAC providers spend nearly \$39 billion a year solely on the administrative activities related to regulatory compliance in these nine domains.

An average-sized community hospital (161 beds) spends nearly \$7.6 million annually on administrative activities to support compliance with the reviewed federal regulations – that figure rises to \$9 million for those hospitals with PAC beds. Nationally, this equates to \$38.6 billion each year to comply with the administrative aspects of regulatory compliance in just these nine domains. Looked at in another way, regulatory burden costs \$1,200 every time a patient is admitted to a hospital.

3. An average size hospital dedicates 59 FTEs to regulatory compliance, over one-quarter of which are doctors and nurses.

Physicians, nurses and allied health staff make up more than one-quarter of the full-time equivalents (FTEs) dedicated to regulatory compliance, pulling clinical staff away from patient care responsibilities. While an average size community hospital dedicates 59 FTEs overall, PAC regulations require an additional 8.1 FTEs.

Retrieval Augmented Generation



- Domain experts identify source documents of interest, available as pdf files.
 - Documents in the public domain, no privacy issues
 - This step adds great value, by defining our 'gold standard'
- Create a vectorstore with chunks from these documents. I used FAISS and langchain, with all-MiniLM-L6-v2 embedding.
- Tool can be hosted publicly on Huggingface spaces. I used llama-2 model under the hood, but only chose to return matching chunks for this project, ignoring the answer (adroit assistant philosophy).

Regulation Documents

The tool searches through documents handpicked by our regulatory team:

- CMS State Operations Manual, Appendix A.
- RCWs - State Law

State Operations Manual Appendix A - Survey Protocol, Regulations and Interpretive Guidelines for Hospitals

Table of Contents
(Rev. 216, 07-21-23)

Transmittals for Appendix A

Survey Protocol

RCW 69.45.040 Storage and transportation of drug samples—Disposal of samples which have exceeded their expiration dates. (1) Drug samples shall be stored in compliance with the requirements of federal and state laws, rules, and regulations.

Certified on 9/1/2023

Page 3

(2) Drug samples shall be maintained in a locked area to which access is limited to persons authorized by the manufacturer.
(3) Drug samples shall be stored and transported in such a manner as to be free of contamination, deterioration, and adulteration.
(4) Drug samples shall be stored under conditions of temperature, light, moisture, and ventilation so as to meet the label instructions for each drug.
(5) Drug samples which have exceeded the expiration date shall be physically separated from other drug samples until disposed of or returned to the manufacturer. [1987 c 411 § 4.]

Revised Code of Washington (RCW)

Last Update: September 1, 2023

The Revised Code of Washington (RCW) is the compilation of all permanent laws now in force. It is a collection of Session Laws (enacted by the Legislature, and signed by the Governor, or enacted via the initiative process), arranged by topic, with amendments added and repealed laws removed. It does not include temporary laws such as appropriations acts.

The Statute Law Committee declares that the certified PDF publication documents in the RCW Archive area on the Office of the Code Reviser's website constitute the official publication of the Revised Code of Washington.

WA Hospital Regulatory Chatbot

Spaces | sonali-tamhankar/WA-Hospital-Regulation... like 2 • Running App Files Community Settings

Find regulations for hospitals in the state of Washington.

We look into these sources to find top ten most relevant excerpts:

- [CMS State Operations Manual Appendix A](#)
- [Pharmacists](#)
- [Pharmacy Assistants](#)
- [Regulation of Health Professions—Uniform Disciplinary Act](#)
- [Abuse of Children and Adult Dependent Persons](#)
- [Administrative Procedures Act](#)
- [Public Disclosure](#)
- [Department of Health](#)
- [Uniform Food, Drug and Cosmetic Act](#)

Hospital Regulation Chat

Ask your question and [click 'Find excerpts'](#).

e.g. What are the rules regarding a Quality Improvement, or QAPI program?

How should expired medications be handled?

[Find excerpts](#)

CMS State Operations Manual Appendix A

Page: 312

...

- Inspect patient -specific and floor stock medications to identify expired, mislabeled or unusable medications.

CMS State Operations Manual Appendix A

Page: 312

...

- Inspect patient -specific and floor stock medications to identify expired, mislabeled or unusable medications.

A-0506 (Rev. 37, Issued: 10-17-08; Effective/Implementation Date: 10-17-08)

\$482.25(b)(4) - When a pharmacist is not available, drugs and biologicals must be removed from the pharmacy or storage area only by personnel designated in the policies of the medical staff and pharmaceutical service, in accordance with Federal

Access to prescription drugs

Page: 2

...

remain intact; (b) The prescription drug bears an expiration date that is more than six months after the date the prescription drug was donated; (c) The prescription drug or supplies are inspected before the prescription drug or supplies are dispensed by a pharmacist employed by or under contract with the pharmacy, and the pharmacist determines that the prescription drug or supplies are not adulterated or misbranded; (d) The prescription drug or supplies are prescribed by a

...

Drug Samples

Page: 3

...

is beneficial to the citizens of this state." [1989 c 164 § 1.] RCW 69.45.060 Disposal of surplus, outdated, or damaged drug samples. Surplus, outdated, or damaged drug samples shall be disposed of as follows: (1) Returned to the manufacturer; or (2) Witnessed destruction by such means as to assure that the drug cannot be retrieved. However, controlled substances shall be returned to the manufacturer or disposed of in accordance with rules



Generative AI and Machine Learning in Healthcare



Non-profit Healthcare Ecosystem

Regulatory Chatbot



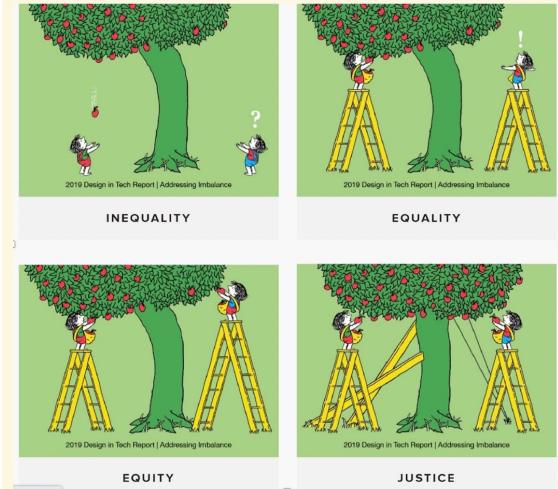
Sickle Cell Emergency Visits

Future Directions



Motivation

- “Inequality, Equality, Equity, and Justice”
- Data scientists can make an impact through our choice of problems, data, methods and tools.
- “The majority of Sickle Cell Anemia (86.3%) and trait (80.2%) cases were among children who were black. Children with SCA were born in 23% of Michigan counties” - Reeves, 2019
- Almost all people with sickle cell disease have painful episodes called crises. These can last from hours to days. Crises can cause pain in the lower back, leg, joints, and chest.
- Some people have one episode every few years. Others have many episodes each year. The crises can be severe enough to require a hospital stay.

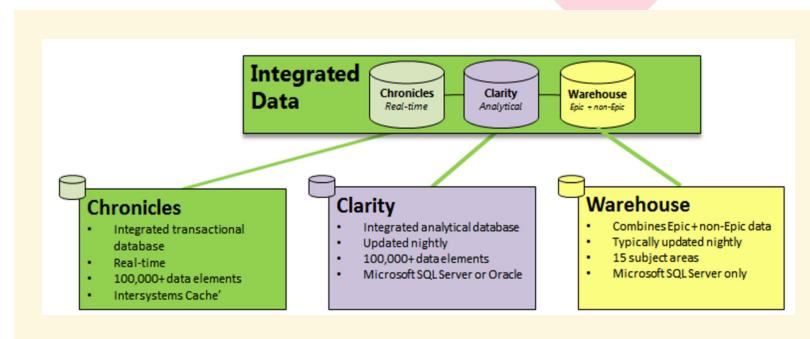


Project Goal

- Sickle cell patients are often seen in the emergency room, especially for sickle cell pain crisis and vaso-occlusive crises (VOC).
- The drug called L-glutamine oral powder (Endari) has proven to help prevent these crises from occurring and thus preventing hospitalizations. Hydroxyurea (Droxia, Hydrea, Silkos) and voxelotor (Oxbryta) prevent abnormal red blood cells from forming. This reduces the number of painful crises from sickling blood cells - WedMD
- Up to a third of these visits are potentially preventable- Nimmer 2015
- Question: Which patients are at highest risk? How can we empower care providers to understand drivers of risk and design effective interventions to reduce avoidable ED visits?
- Goal: Create a Sickle Cell ED Visits Tool.
 - Traditional Machine Learning: Build a predictive model for ED risk
 - Generative AI: Augment traditional toolkit for deeper investigations

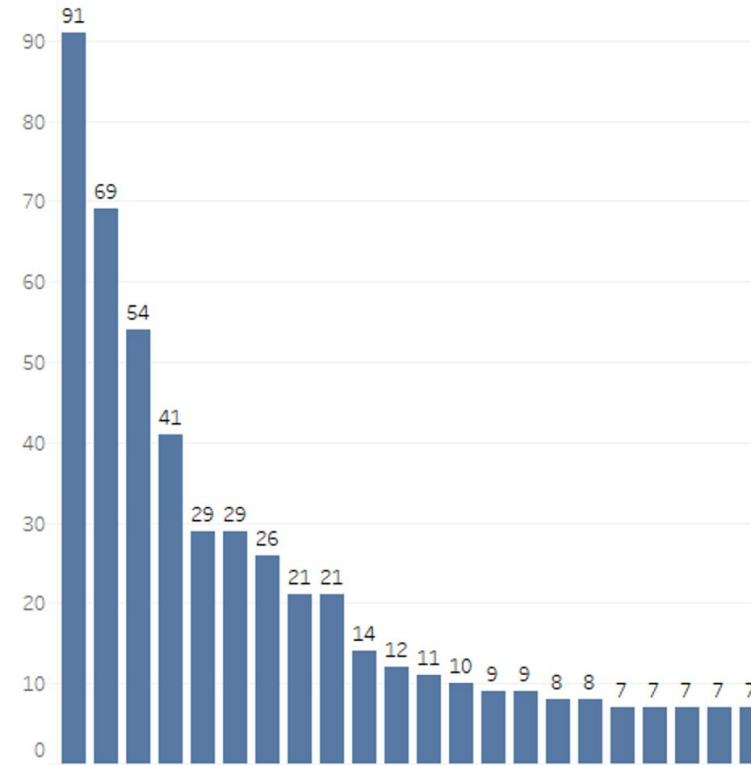
Data

- Define a cutoff date for training model.
- Patients with a diagnosis of sickle cell, exclude those with only sickle cell trait, or with just a family history.
- Retain patients with at least two visits in the year before cutoff date.
- Features (one year before cutoff)
 - Demographic
 - Clinical
 - Medications
 - Labs
 - Operational
 - Past encounters, reason for visit
 - Past emergency room visits
- Target variable - ED visit in one month after cutoff date

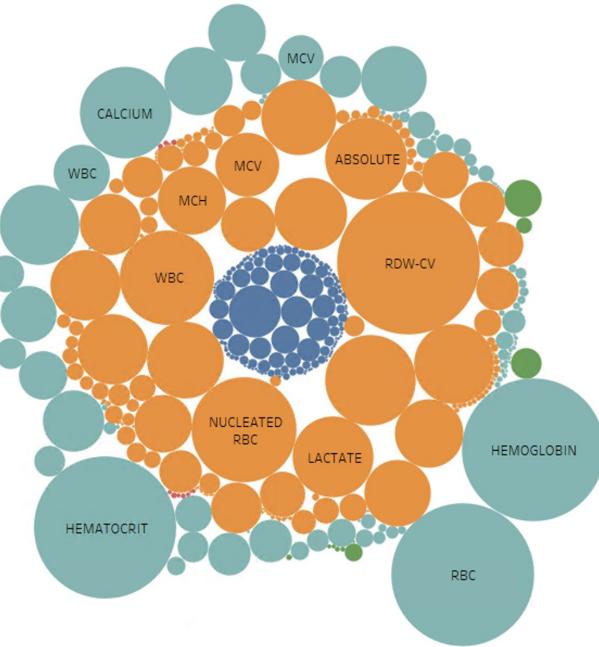
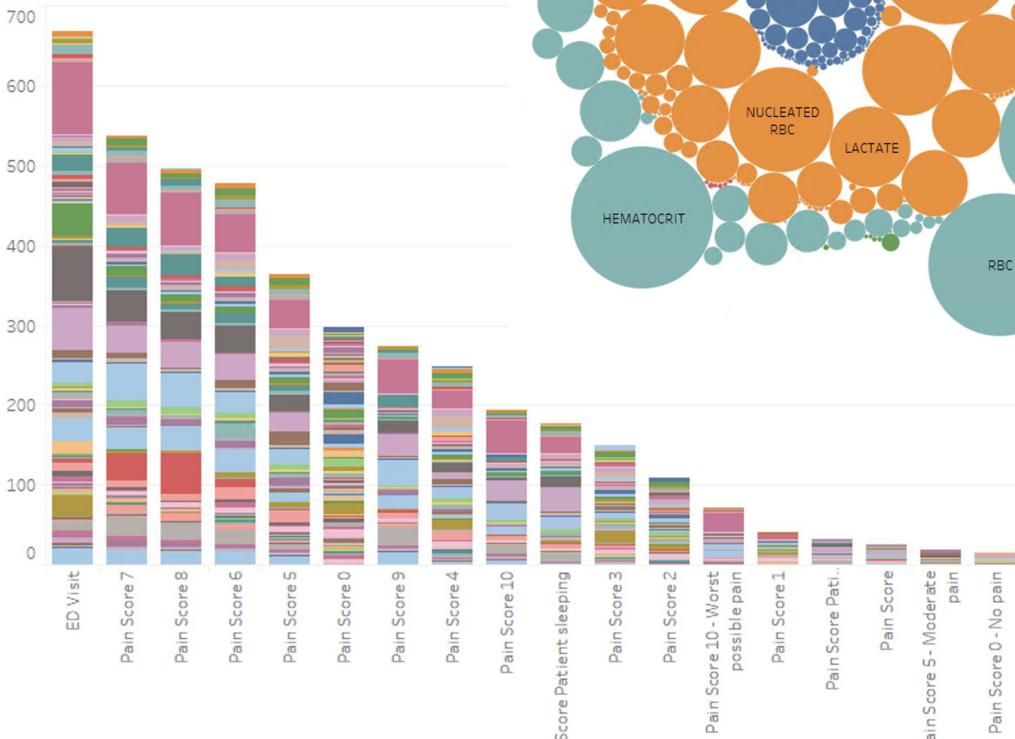
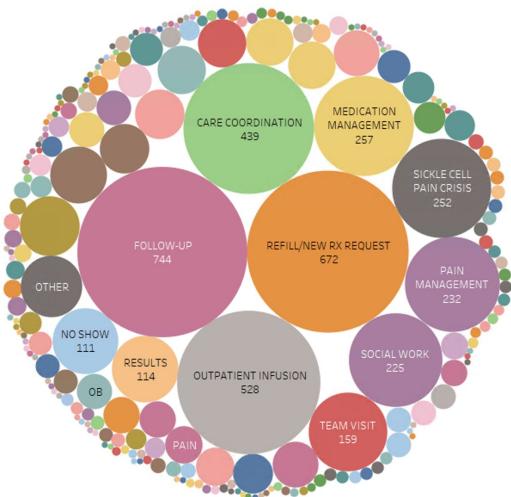


Exploratory Analysis

- 136 patients qualify in the denominator.
- About a dozen of these patients have over 12 emergency room visits in the past year.
- Data summary:
 - Average of 0.67 ED visits per patient in the year before
 - 81 women, 55 men
 - 120 out of 136 patients are Black or African American
 - Mean age = 34 years



Exploratory Analysis

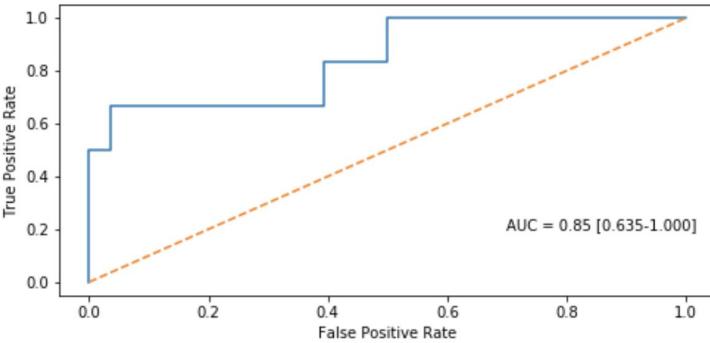
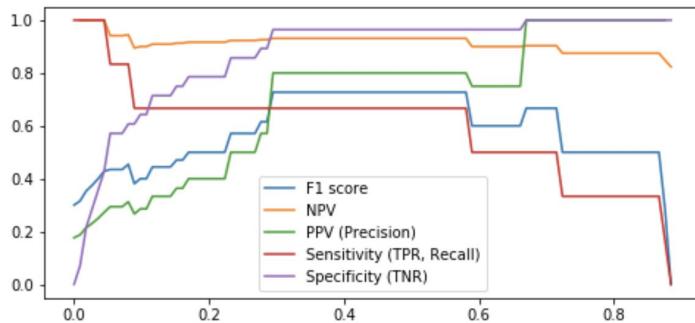


Model

- 136 patients, 23 had an ED visit within a month (target)
- Medications, lab values, reason for visit:
 - remove infrequent values
 - pivot to count the number of times a value appears for a patient
 - labs: abbreviation_flag = abbreviation for lab, is it low/high/abnormal
 - a total of 372 columns
- keep top 50 with highest correlation with the target
- 75-25 train-test split
- Random forest model
 - n_estimators = 200
 - max_depth = 6
 - max_leaf_nodes = 9

Area under the ROC curve = 0.85

Credit: Penn Medicine



importance	var
0.102123	reason_visit_name_pain_management
0.090327	name_enoxaparin_sodium_40_mg_0_4ml_ij_sosy
0.081873	reason_visit_name_no_show
0.078147	name_ketorolac_tromethamine_30_mg_ml_ij_soln
0.061249	name_naloxone_hcl_0_4_mg_ml_ij_soln

How do we make these findings actionable?

We have a model.
Now what?

How do we make these findings actionable?

7 out of 10 physicians
feel overwhelmed by
the amount of
information

We have a model.
Now what?

How do we make these findings actionable?

7 out of 10 physicians
feel overwhelmed by
the amount of
information

Shapley Values

“The core idea behind Shapley value based explanations of machine learning models is to use fair allocation results from cooperative game theory to allocate credit for a model’s output among its input features.”

We have a model.
Now what?

How do we make these findings actionable?

7 out of 10 physicians
feel overwhelmed by
the amount of
information

Shapley Values

“The core idea behind Shapley value based explanations of machine learning models is to use fair allocation results from cooperative game theory to allocate credit for a model’s output among its input features.”

We have a model.
Now what?

A large percent (80% - HIT Consultant, 2015) of healthcare data is unstructured.

How do we make these findings actionable?

7 out of 10 physicians feel overwhelmed by the amount of information

Shapley Values

"The core idea behind Shapley value based explanations of machine learning models is to use fair allocation results from cooperative game theory to allocate credit for a model's output among its input features."

We have a model.
Now what?

A large percent (80% - HIT Consultant, 2015) of healthcare data is unstructured.

Generative AI

Tools based on large language models can allow clinicians to converse with patient notes in natural language! Can't afford hallucinations - use these tools as adroit assistants, with easy reference to sources.

Sickle Cell ED Visits Tool

The previous project was hosted publicly on huggingface. Now, we want to work with protected health information.

We want to run models locally for now.

- currently running on AWS g5.2xlarge instance
- using ollama for easy hosting of llama2 model
- web interface based on streamlit

Sickle Cell Tool Mock-up

Enter MRN:

Med list for patient

List of Highest Risk MRNs for ED Visit in the next 6 months

ED Visit Dates in the last 6 months

Patient Risk and Risk factors plot

Chat against patient notes Pull exact notes from a date

Sickle Cell ED Visits Tool

- The tool is password protected, hosted within organizational firewall.
- After validating credentials, a list of highest risk patients is displayed in sidebar.
- Specific patient MRN can be entered for further investigation.



Highest risk patients

Patient MRN	Patient Name
PATMRN1	John Doe
PATMRN2	Jane Doe
PATMRN3	James Doe
PATMRN4	Jeffrey Doe
PATMRN5	Jenny Doe
PATMRN6	Jasmine Doe

Sickle Cell ED Visits Tool

Access to protected health information on this page is limited to authorized users.

Enter your username.

Enter your password



Enter the MRN.

Sickle Cell ED Visits Tool

Pain Score 07

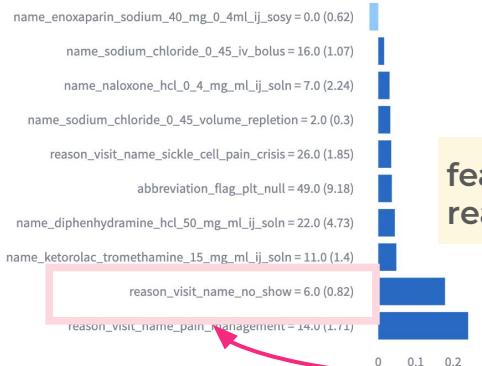
ED Visit

Dates

Pain Scores and ED Visits



Features Most Impacting ED Risk



feature = patient value(cohort mean)
reason_visit_name_no_show = 6.0(0.82)

Enter your question here, for an answer based on clinical notes and references.

Submit

“Enter your question here, for an answer based on clinical notes and references.”

Sickle Cell ED Visits Tool

Two tabs:
“Relevant Note Excerpts” and
“Clinical Notes Full Text”.

Enter your question here, for an answer based on clinical notes and references.

Why did this patient miss appointments? How many appointments did they miss?

Submit

[Relevant Note Excerpts](#) [Clinical Notes Full Text](#)

Based on the information provided, the patient missed two appointments. The first appointment was missed due to transportation issues and the patient could not make it to the second appointment because he was in pain "like labor pain that did not stop".

...

Here are excerpts from patient notes:

2023-05-23

The patient did not arrive for his scheduled RN visit today. This visit was to complete clinic orientation and paperwork as well as a pain assessment with medication refill. The patient was initially scheduled for this visit on April 18th and

...

2023-08-27

John was scheduled to see Dr. Whitecoat today, but did not show up for his appointment. On follow-up call, he reported that he was in bed all day due to pain, slept most of the day and completely forgot about the appointment. He reported that his pain felt like labor pain that did not stop.

“Why did this patient miss appointments?
How many appointments did they miss?”

I do display the answer so we can validate and test its quality and usefulness.

Relevant note excerpts are pulled from this patient’s notes.
Retrieval Augmented Generation is used under the hood, with patient MRN and visit date stored as metadata for chunks.

Sickle Cell ED Visits Tool

Enter your question here, for an answer based on clinical notes and references.

Why did this patient miss appointments? How many appointments did they miss?

Submit

Relevant Note Excerpts [Clinical Notes Full Text](#)

Encounter Date:

2023/01/01

Second tab can provide the verbatim note text for a specific date, so providers can investigate further.

Generative AI and Machine Learning in Healthcare



Non-profit Healthcare Ecosystem

Regulatory Chatbot



Sickle Cell Emergency Visits

Future Directions



Work in Progress

- Improve Sickle Cell Tool
 - Validate and refine
 - Model improvements
 - Interface enhancements
- Widen scope to ED Visits for all patients
 - The National Institutes of Health's Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity (AIM-AHEAD) fellowship stakeholder project
 - Larger cohort, more diverse population
- Technical questions
 - Infrastructure considerations
 - current setup will not scale for larger vectorstores
 - Larger LLMs could be considered for better quality answers
 - More sophisticated responses using agents to pull different aspects of patient record

Takeaways

- Healthcare ecosystem is complex, especially in a care delivery organization. There are unique opportunities for meaningful work, if we can creatively address the challenges of working in this space.
- Each of us has the power to make a difference through our daily work!
- Partnering with domain experts is critical, their input is hard to replace with automated extractions.
- Tools based on large language models can add significant value, especially when used responsibly as adroit assistants to users who can easily access source information.
- Combining traditional machine learning toolkit with generative AI can help us design interactive tools for actionable insights.
- Enormous opportunities exist in this vast and fast moving space with powerful developer tools.
- Thankful to the Data Science Salon for this knowledge sharing and networking opportunity - let's connect!

Thank You!