

Extending the capabilities and reach of EMERSE in support of cancer research


David Hanauer
University of Michigan

November 3, 2023

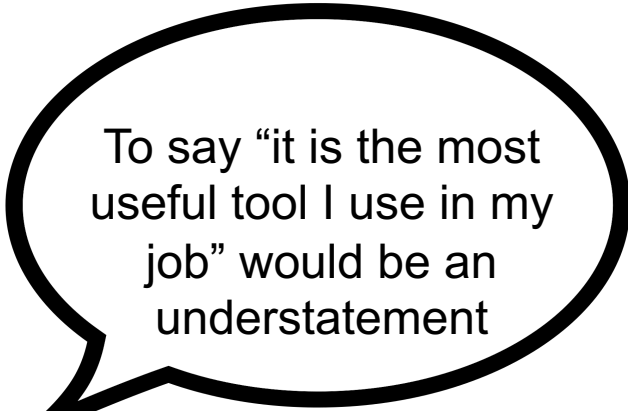


THANK YOU
to ITCR
for the generous support!


And thank you on behalf of our users



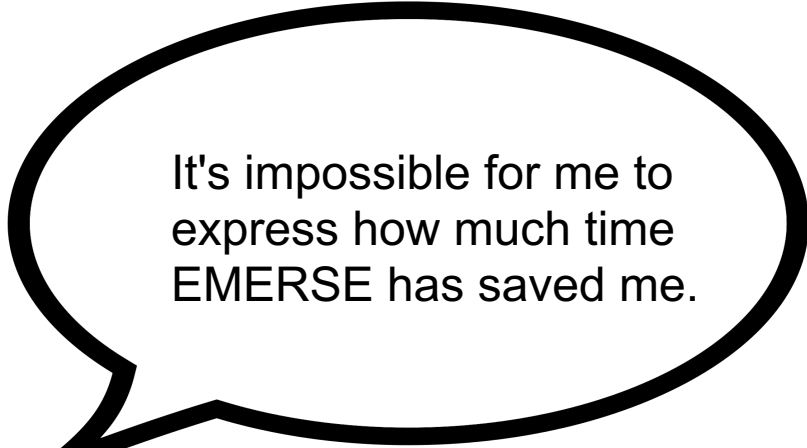
Thanks so much,
EMERSE is
SOOOO important
in our daily work!



To say “it is the most
useful tool I use in my
job” would be an
understatement



We are truly
thankful for a tool
like Emerse



It's impossible for me to
express how much time
EMERSE has saved me.



What is EMERSE

- Information retrieval tool to support clinical research using clinical notes from electronic health records (EHRs)
- Enterprise-level software (one installation per site)
- Designed for non-technical researchers to easily access and utilize the notes
- Designed for system administrators with security in mind
- Supports tasks important to clinical research
 - Cohort identification
 - Data abstraction
 - Patient Lists

Why is EMERSE needed?



- ~80% of clinical data are in free text form
- NLP tools remain difficult to use
- Data security can no longer be an afterthought

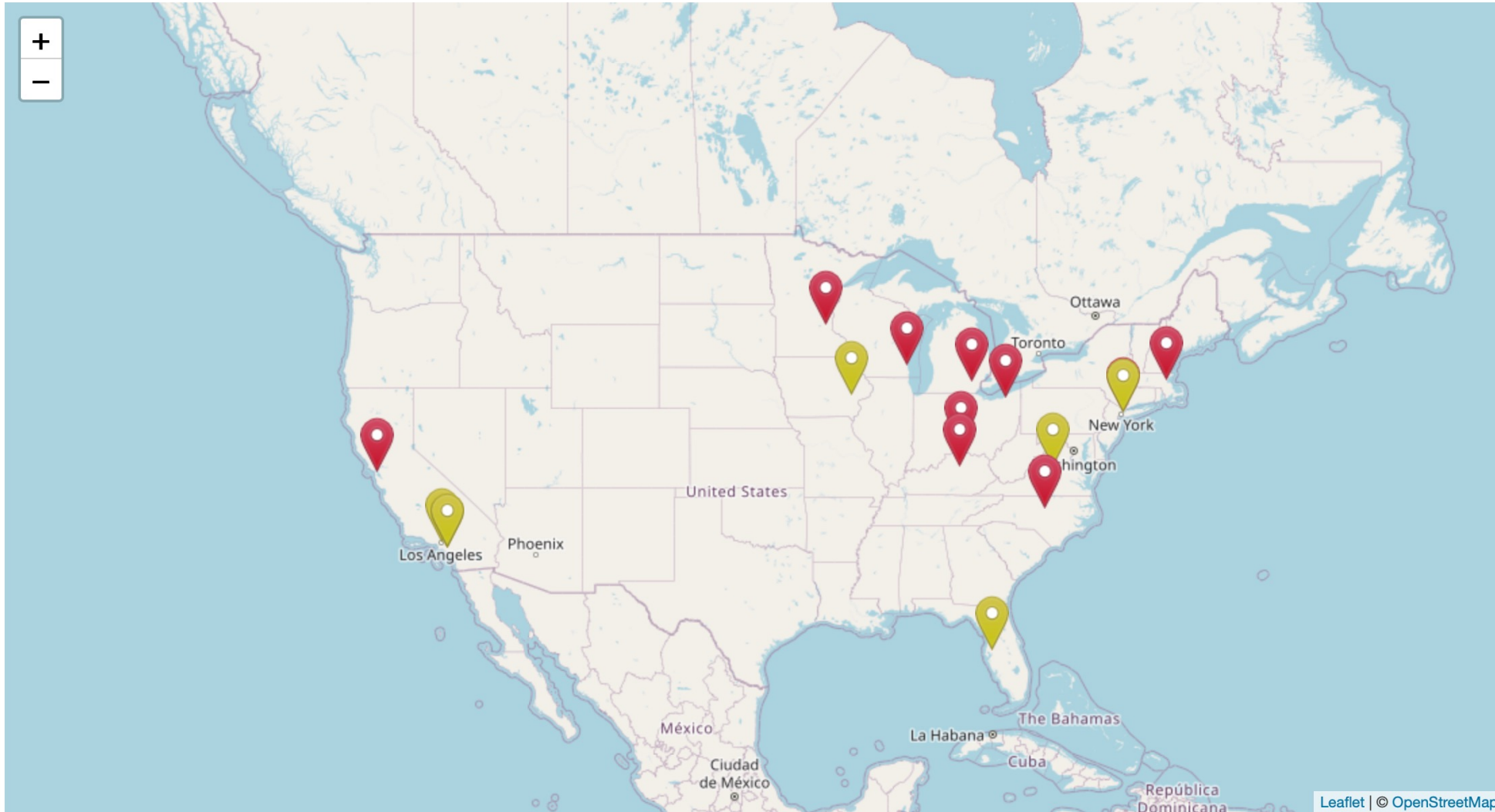
Where is EMERSE – a growing community

Locations of EMERSE sites, installed:  and installing: 



Where is EMERSE – a growing community

Locations of EMERSE sites, installed:  and installing: 



Recent publications using EMERSE

- Impact of Metagenomic Next-Generation Sequencing of Plasma Cell-free DNA Testing in the Management of Patients With Suspected Infectious Diseases
- Low-parameter supervised learning models can discriminate pseudoprogression and true progression in non-perfusion-based MRI
- Margin status in vulvovaginal melanoma: Management and oncologic outcomes of 50 cases
- Subcutaneous Leiomyosarcoma: An Aggressive Malignancy Portending a Significant Risk of Metastasis and Death
- Demographic and clinical characteristics of patients with metastatic breast cancer and leptomeningeal disease: A single center retrospective cohort study
- Social Determinants of Health and Racial Disparities in Cardiac Events in Breast Cancer
- Social Determinants of Health Data Improve the Prediction of Cardiac Outcomes in Females with Breast Cancer
- Thirty-Day Unplanned Hospital Readmissions in Patients With Cancer and the Impact of Social Determinants of Health: A Machine Learning Approach
- The Development of Head and Neck Cancer in Patients with the Isolated Complaint of Globus Pharyngeus

Proposed work in 4 themes

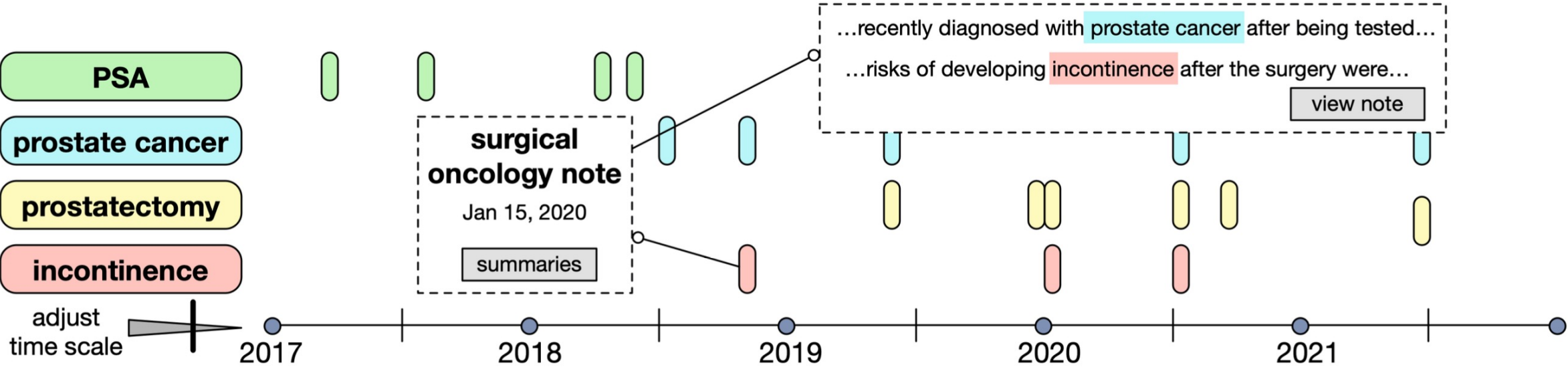
- New Features
- Natural Language Processing (NLP) integration
- Network
- Evaluation

New Features

- Standards-Compliant API
 - Needs to respect roles/privileges from the application
 - Detailed audit logs
 - Support technical use cases (finding documents, cohorts, patient lists, etc)
 - Test with ITCR DeepPhe (Cancer Deep Phenotyping) integration

New Features

- Timeline data visualizations



New Features

- Data Extraction Workflow Editor

The interface is divided into three main sections:

- Data Extraction Pathway Items:** A list of 13 items to be dragged into the editor:
 - IDENTIFY NOTE THAT CONTAINS
 - SELECT LINE THAT STARTS WITH
 - SELECT LINE THAT CONTAINS
 - REMOVE TEXT AFTER, BUT NOT INCLUDING
 - REMOVE REGULAR EXPRESSION
 - CLEAN TEXT - TRIM
 - SELECT TABLE COLUMN THAT CONTAINS
 - SELECT TABLE ROW THAT CONTAINS
 - MAP TEXT
 - ADD RESULT TO COLUMN
- Data Extraction Pathway Editor:** A workspace for building a pathway. It contains:
 - Instructions: "The Pathway will be run in the order from top to bottom. The first item must be 'identify note that contain'. The last item must be 'add result to column'. The order of the items can be rearranged by dragging them."
 - Items in the pathway:
 - IDENTIFY NOTE THAT CONTAINS: Research Note v1.53
 - SELECT LINE THAT STARTS WITH: stage:
 - EXTRACT EVERYTHING AFTER THE TEXT: stage: (with edit icon)
 - REMOVE TEXT FROM LINE: (AJCC system) (with checkbox for case-sensitive and Done button)
 - CLEAN TEXT - TRIM
 - MAP DATA (with Select map button)
 - ADD RESULT TO COLUMN: Cancer Stage
 - Buttons: Clear pathway, Add Pathway to Workflow
- Data Extraction Workflow Editor:** A workspace for managing a workflow. It contains:
 - Instructions: "Each Pathway below will extract data to a single column in the spreadsheet. The order of the columns will match the order of the pathways below. The order of the Pathways can be rearranged by dragging them."
 - More Pathways... button (with callout: "pre-built pathways can be selected here")
 - Pathways in the workflow:
 - MRN (with Remove button)
 - Diagnosis (with edit icon)
 - PHQ-9 (with edit icon)
 - Body Mass Index (with Edit and Remove buttons)
 - Warning dialog: "!! There is another pathway being edited. Are you sure you want to replace it with this one?" (with Cancel and OK buttons)
 - Save Workflow button

New Features

- Data Extraction Workflow Editor

Research Note v1.53

Medical record number: 3532352

Active diagnosis: breast ca

Stage: IIA (AJCC system)

History: Mrs. Jones is a 62-year-old female who was seen in our clinic for a consultation. Depression screening was done today.

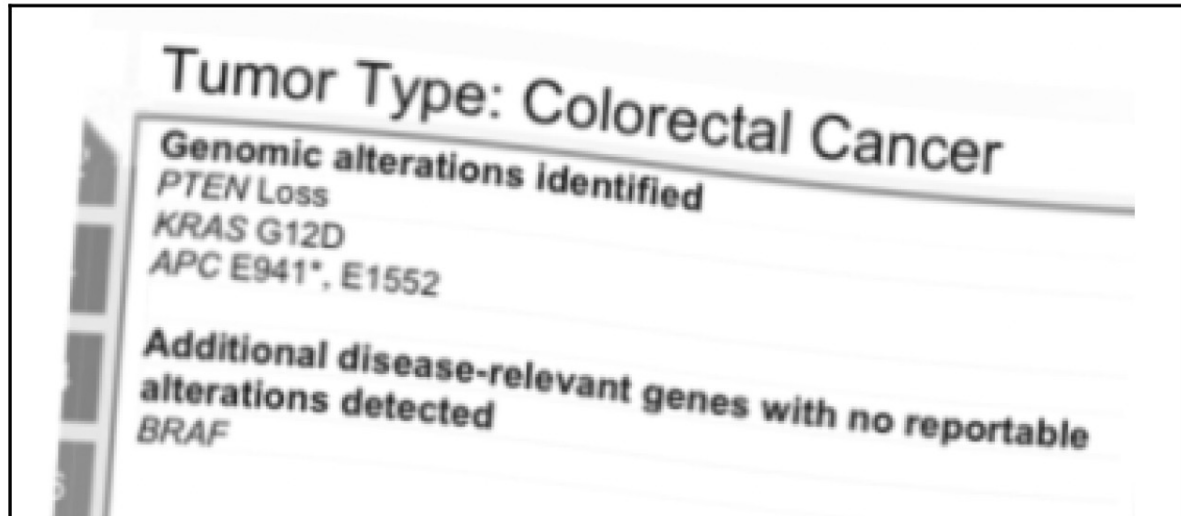
PHQ-9 score: 6

Physical exam: **weight:** 80 kg, **height:** 170 cm

MRN	Diagnosis	PHQ-9	Body Mass Index	Cancer Stage
3532352	breast ca	6	27.63	IIA
3235655	breast ca	8	22.12	IIB
7547853	breast ca	2	18.32	IB
8553334	breast ca		29.23	Missing
2237754	breast ca	3	20.43	IIB

New Features

- OCR Pipeline

 <p>Tumor Type: Colorectal Cancer</p> <p>Genomic alterations identified</p> <p>PTEN Loss</p> <p>KRAS G12D</p> <p>APC E941°, E1552</p> <p>Additional disease-relevant genes with no reportable alterations detected</p> <p>BRAF</p>	<p>Tumor Type: Colorectal Cancer</p> <p>Genomic alterations identified</p> <p>PTEN Loss</p> <p>KRAS G12D</p> <p>APC E941°, E1552</p> <p>Additional disease-relevant genes with no reportable alterations detected</p> <p>BRAF</p>
---	---

NLP

- Incorporate Natural Language Processing (NLP) into EMERSE
 - “Out-of-the-Box” capabilities
 - Named Entity Recognition (mapped to UMLS CUIs & semantic types)
 - Negation (***) most requested feature by users)
 - Uncertainty
 - Subject
 - History of

NLP

- Build generalizable/extensible NLP import to accommodate external solutions, test with 3 partner institutions
 - UCSF
 - Moffitt
 - City of Hope
- Assumption is that others may want to replace or augment our “out-of-the-box” NLP
- Will help sites keep up with latest technologies/capabilities
- In later years, partner with 2 NLP expert teams to make sure EMERSE is keeping up with latest advancements: cTAKES & MedTagger teams

▼ NLP Details

Anatomy (81) Attribute (0) Device (0) Disorder (114) Drug (68) Event (0) Finding (504) Lab (0)
Procedure (301) Subject (0) Title (0) All Groups (1068)

Negation (15)		Uncertainty (17)
Non-patient subject (7)		

Summary:

...further investigating the use of chemotherapy in pregnancy. Case presentation FOLFOX-6 was administered...

Document:

Case Report from the Journal of Medical Case Reports Title: Chemotherapeutic treatment of colorectal cancer in pregnancy: case report | Journal of Medical Case Reports | Full Text
DOI: 10.1186/s13256-015-0621-9 Abstract Introduction Colon cancer in pregnancy is uncommon. Only a small number of case reports have been published in the literature on the use of chemotherapeutic drugs during pregnancy. Reports of such cases assist clinicians in further investigating the use of chemotherapy in pregnancy. Case presentation FOLFOX-6 was administered to a pregnant, 33-year-old Saudi woman with metastatic colon cancer. She tolerated the chemotherapy well and delivered a full-term baby girl with no obvious harm. Colon cancer during pregnancy is not easily detected and is difficult to manage. A detection suggestive of malignancy. A multidisciplinary approach with patient involvement and to limit side effects for the fetus. Further data and long-term follow-up are needed for offspring. Open Peer Review reports Introduction Colorectal cancer (CRC) in pregnancy. Symptoms of CRCs overlap with expected physiological changes in pregnancy and associated issues, and limited literature exists to guide a management approach. Over 300 cases of CRC use during pregnancy have been reported [3-6], with five additional cases reported from Camden, NJ, USA [7]. We report a case of a 33-year-old pregnant woman with metastatic rectal cancer treated with chemotherapy with no apparent fetal harm. Case presentation A 33-year-old Saudi woman (gravida 9, para 8) presented to our emergency department at 11 weeks of gestation with abdominal pain and increased flatus. The patient had a history of bleeding per rectum for 21-2 years, but she believed it was due to hemorrhoids and failed to mention it to her primary care physician. Abdominal ultrasound revealed a complex left ovarian mass (19*12cm) extending up to the epigastrium with hyperechoic solid components, and a small amount of fluid was seen in the pelvis surrounding the appendix. Her

Disorder

- C0346975: metastatic rectal cancer
- C0861772: metastatic rectal cancer
- C00071113: rectal cancer
- C0006826: cancer

Finding

- C0346975: metastatic rectal cancer
- C0861772: metastatic rectal cancer

× cancer was diagnosed during her pregnancy. She tolerated chemotherapy well and delivered a full-term baby girl with no obvious harm at her 2-year follow-up examination. Conclusion Colon cancer during pregnancy is not easily detected and is difficult to manage. A multidisciplinary approach with patient involvement and to limit side effects for the fetus. Further data and long-term follow-up are needed for offspring. Open Peer Review reports Introduction Colorectal cancer (CRC) in pregnancy. Symptoms of CRCs overlap with expected physiological changes in pregnancy and associated issues, and limited literature exists to guide a management approach. Over 300 cases of CRC use during pregnancy have been reported [3-6], with five additional cases reported from Camden, NJ, USA [7]. We report a case of a 33-year-old pregnant woman with metastatic rectal cancer treated with chemotherapy with no apparent fetal harm. Case presentation A 33-year-old Saudi woman (gravida 9, para 8) presented to our emergency department at 11 weeks of gestation with abdominal pain and increased flatus. The patient had a history of bleeding per rectum for 21-2 years, but she believed it was due to hemorrhoids and failed to mention it to her primary care physician. Abdominal ultrasound revealed a complex left ovarian mass (19*12cm) extending up to the epigastrium with hyperechoic solid components, and a small amount of fluid was seen in the pelvis surrounding the appendix. Her

Anticipated future side project: LLMs

- Already exploring possibilities for integration and use cases
 - Will be testing accuracy, effectiveness, usefulness, cost, etc.
 - Consider security, privacy of protected health information
 - Local versus external/commercial models
-
- Grant deadline for our submission cycle: November 17, 2022
 - Release date of OpenAI ChatGPT 3.5: November 30, 2022

Network

- EMERSE has a built-in network capability to obtain obfuscated patient counts across sites
 - Useful for cohort estimation, especially for rare disorders
- Sites have been reluctant to turn it on, due to security concerns
 - Despite multiple safeguards built-in
 - Security was baked into the design
- Plan to work with 4 sites to develop a generalizable network participation framework, involving: security, regulatory, and other groups to understand and address all concerns

Network

- Incorporate data standards to enable better network search filtering
 - Few standards exist for unstructured data
 - We can start with “simple” mapping for demographics
 - Will consider various standards including OMOP CDM
- Develop advanced auditing to ensure queries are not attempting to re-identify individual patients
- Add ability to exclude “special status” patients from local or networked searched.

Evaluation



Online User Surveys



Site Lead Interviews



Scientific/Clinical Outcomes

Miscellaneous

- Many details can be found at: <https://project-emerse.org>
- Contact us at: EMERSE-team@umich.edu
- Thoughts on set-aside collaborations? Contact us!
- EMERSE Community Meetings 3-4 times/year, next in Jan/Feb 2024
- EMERSE Research Collaborative, meeting ~monthly
- Video recordings of prior presentations at <https://project-emerse.org/presentations.html>
- Sneak preview of NLP integration: <https://vimeo.com/873431543>