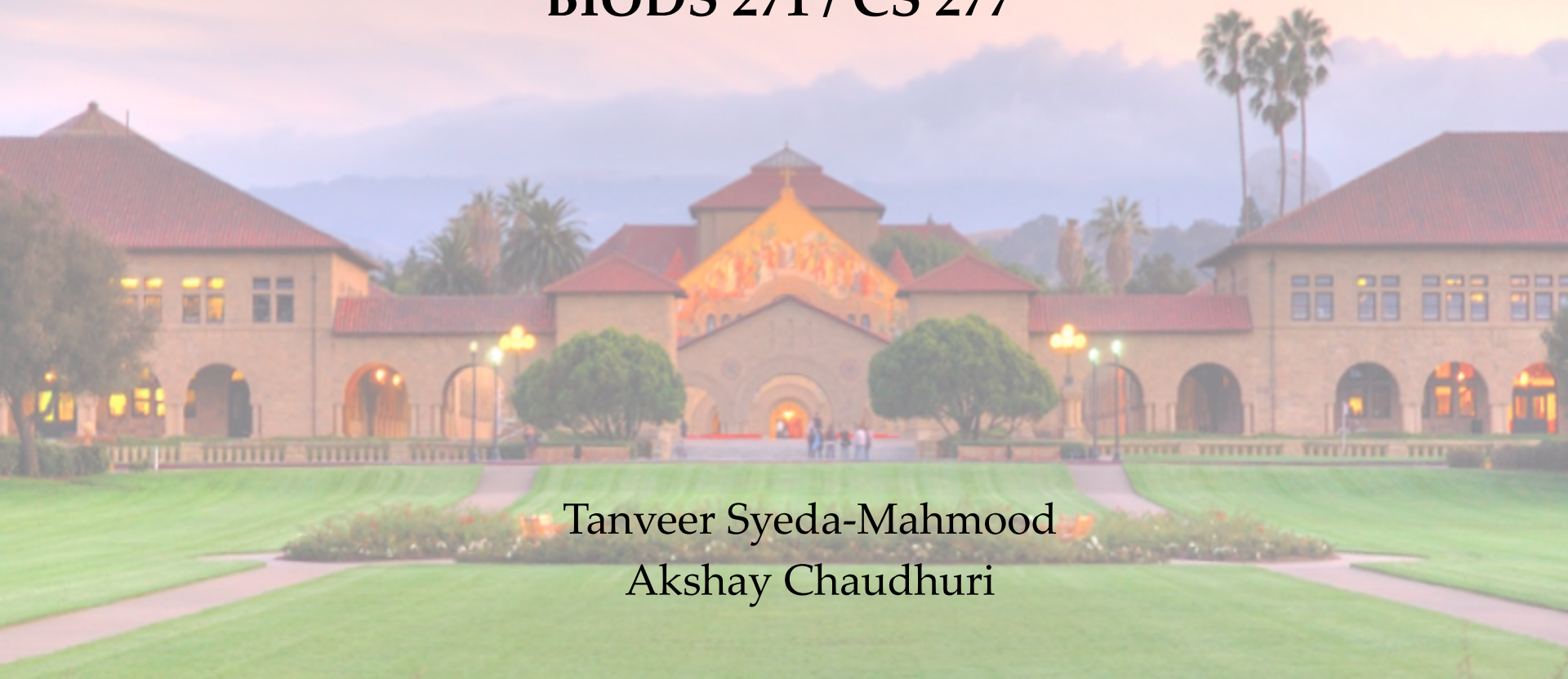


# Clinical Needs for Foundational Models

BIODS 271 / CS 277



Tanveer Syeda-Mahmood  
Akshay Chaudhuri

# Healthcare-specific Foundation Models

- Some target tasks addressed by foundational models may not be as useful for healthcare-specific use cases:
  - *Sentiment analysis*
  - *Masked language prediction*
  - *Question answering*
  - *Captioning*
  - Language translation
  - Summarization
  - Language generation
  - Classification
  - Segmentation
  - Speech to text

# User base for Healthcare FM

- Providers/Staff/Technicians
  - Clinicians: : Primary care, PCP, Radiologists, Cardiologists, Oncologists, Radiation oncologists, Interventional cardiologists, radiologists, ...
  - Nursing and other hospital staff
  - Technicians
- Payers/Backoffice
  - Insurance, payer-provider networks, Medicare
  - Backoffice staff
- CIO/IT Staff
  - IT/Data warehousing, CRO
  - Training
- Clinical researchers/Academics
  - Clinical study research
  - Clinical education & training
- Patients

# Clinical Needs for FMs

## Imaging



- Technical assessment during imaging
- Worklist prioritization
- Referral routing, Telemedicine
- **Preliminary read – disruptive technology to expedite a dictation-driven workflow**

X-ray Read Service



Imaging study

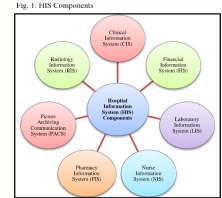
Preliminary report

The mediastinum is prominent and there is also cardiomegaly. (Key) associated by low lung volumes and vertical positioning.  
Possible background hyperinflation/atelectasis.  
Effusions would be difficult to exclude.  
Enlarged cardiac mediastinal silhouette with various aorta.  
There is upper airway retraction and no airway dilation/engorgement.

Radiology Read



EMR



**FINDINGS:**  
**BRAIN PARENCHYMA:**  
T2 hyperintense white matter lesions:  
\*Periventricular: [Six]: (2) or more) lesions contacting the ependymal surface  
\*Juxtacortical/Cortical: [None]/[Present]  
\*Infratentorial: [None]/[Present] involving the [brainstem/cerebellum/brainstem and cerebellum]  
\*Optic Nerves: [None]/[Present]  
\*Cervicomedullary junction: [None]/[Present]  
**Enhancing Lesions:** (4 of enhancing lesions and locations=None)  
**Reduced Diffusion:** [None]/[Present] (described)  
**Overall Disease Burden:** [None]<10 lesions|10-20 lesions|>20 lesions)  
**Parenchymal Atrophy:** [None]/[Mild/Moderate/Severe]  
**Callosal Atrophy:** [None]/[Mild/Moderate/Severe]  
**OTHER FINDINGS:** [None]  
**IMPRESSION:**  
[Normal MRI brain]. Multiple white matter lesions that are [typical/atypical/not consistent] with demyelinating disease(s).  
  
2016 MAGNIMS MRI criteria to establish disease dissemination in space in multiple sclerosis (Lancet Neurol. 2016; Mar;15(3):292-303)  
Involvement of at least two of five areas of the CNS as follows:  
\* Three or more periventricular lesions  
\* One or more infratentorial lesions  
\* One or more spinal cord lesions  
\* One or more optic nerve lesions  
\* One or more cortical or juxtacortical lesions

Radiologist-corrected report

# What are the clinical needs of providers?

- Technical assessments of imaging exams
- Screening and triage (first reader)
- Disease segmentation and tracking
- Measurement quantification
- Patient summaries, KPI
- Diagnostic decision support (second reader)
  - Recommendations for diagnosis, treatment or outcome
  - Normal/abnormal discrimination
  - Anomaly detection & classification
- Automated reporting
- Foundational Models
  - ✓ DNN
  - ✓ LM
  - ✓ LLM
  - ✓ VLM
- FM Workflows
  - ✓ Applications using multiple FM

# Clinical text analysis

MR #: XXXXXXXX  
 Patient Location: OP  
 Blood Pressure: /  
 Age: 86  
 Gender: M  
 HT:

Referring MD: XXXXXX  
 Indica:

M-Mode

RVID

LVID

LA

A0

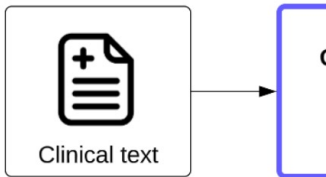
IVS

LVPW

EPSS

FS

LV EF



Study (Rhythm:  
 HR: bpr

Left V  
 ular :  
 grossl  
 n. Over  
 fraction is estimated to be 65 to 70%. There is mi  
 eft ventricular hypertrophy. No regional  
 wall motion abnormalities are seen.

Right Atrium: Right atrial size is mildly dilated.  
 Right Ventricle: Right ventricular cavity size is  
 ickness is normal, and the free wall  
 contraction is normal.  
 Aortic Valve: The aortic valve is restricted, thick  
 ified. Severe aortic stenosis is present.  
 The aortic valve peak velocity is 4.49 m/s, the pe  
 80.6 mmHg, and the mean gradient is  
 46.0 mmHg. The aortic valve area is estimated to b  
 d to moderate aortic valve  
 insufficiency.

Sonographer: X

**Input:** The patient takes coumadin 5 mg daily for a TIA and she has an aspirin allergy.

## Zero-shot prompt:

**Input:** The patient takes coumadin 5 mg [...].  
**Prompt:** Create a list of medications.  
 She takes 5 mg of Coumadin and Aspirin

Complex post-processing (resolver) of LM output → [coumadin, aspirin]

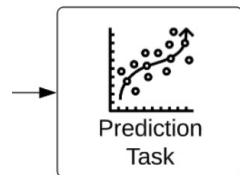
## One-shot example + guidance:

**Input:** He is on a statin now and we think [...].  
**Prompt:** Create a list of medications.  
 -"statin"  
**Input:** The patient takes coumadin 5 mg [...].  
**Prompt:** Create a list of medications.  
 -"coumadin"  
 -"aspirin"

Minimal post-processing (resolver) of LM output → [coumadin, aspirin]

ened, and calc

ik gradient is



	Precision	Recall
	46.7%	79.6%
	72.6%	90.4%
	34.0%	59.6%
	78.2%	79.3%

Echocardiogram  
 collection

753  
 reports

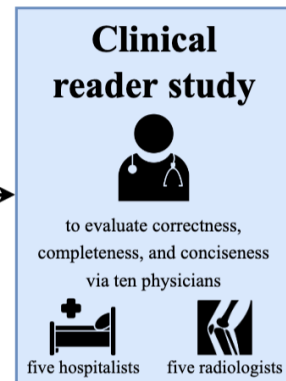
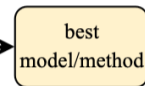
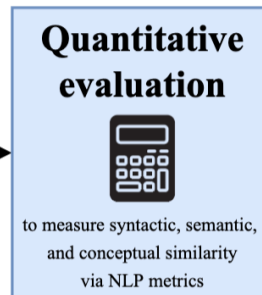
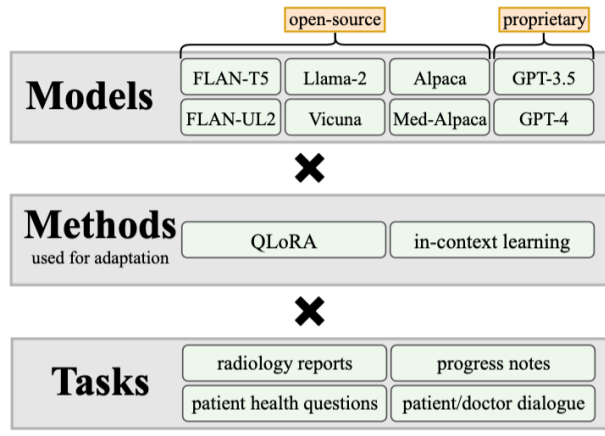
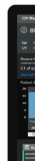
echo

cTakes

WHIA

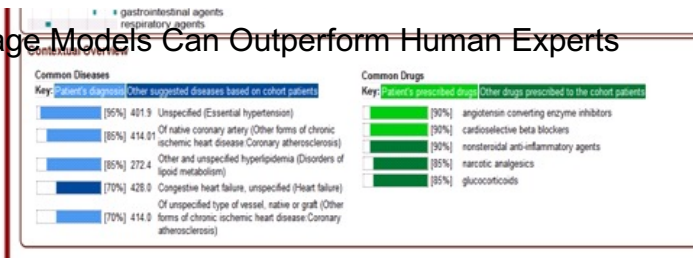
Large Language Models are Few-Shot Clinical Information Extractors, EMLP2022

# Patient summaries



## Clinical Text Summarization: Adapting Large Language Models Can Outperform Human Experts

- reports
- Social and family history
- Demographics
- Vitals
- Allergies
- Labs





# Clinical Decision Support

JAMA  
Network | Open™

Original Investigation | Oncology

## Leveraging Large Language Models for Decision Support in Personalized Oncology

Manuela Benary, PhD; Xing David Wang, MSc; Max Schmidt, MD; Dominik Soll, MD; Georg Hilfenhaus, MD; Mani Nassir, MD; Christian Sigler, MD; Maren Knödler, MD; Ulrich Keller, MD; Dieter Beule, PhD; Ulrich Keilholz, MD; Ulf Leser, PhD; Damian T. Rieke, MD

### Abstract

**IMPORTANCE** Clinical interpretation of complex biomarkers for precision oncology currently requires manual investigations of previous studies and databases. Conversational large language models (LLMs) might be beneficial as automated tools for assisting clinical decision-making.

**OBJECTIVE** To assess performance and define their role using 4 recent LLMs as support tools for precision oncology.

**DESIGN, SETTING, AND PARTICIPANTS** This diagnostic study examined 10 fictional cases of patients with advanced cancer with genetic alterations. Each case was submitted to 4 different LLMs (ChatGPT, Galactica, Perplexity, and BioMedLM) and 1 expert physician to identify personalized treatment options in 2023. Treatment options were masked and presented to a molecular tumor board (MTB), whose members rated the likelihood of a treatment option coming from an LLM on a scale from 0 to 10 (0, extremely unlikely; 10, extremely likely) and decided whether the treatment option was clinically useful.

### Key Points

**Question** Can current conversational large language models (LLMs) be used as a tool for personalized decision-making in precision oncology?

**Findings** In this diagnostic study, treatment option identification from 4 LLMs for 10 fictional patients deviated substantially from expert recommendations. Nevertheless, LLMs correctly identified several important treatment strategies and partly provided reasonable suggestions that were not easily found by experts.

Aalim - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Google

Aalim

Advanced Analytics for Information Management

Global Edition Emerging Technologies

## Study: Google reveals LLM that helps accurately diagnosis complex cases

The tech giant revealed it created an LLM optimized for diagnostic reasoning with the ability to generate a differential diagnosis or assist clinicians in coming to a diagnosis.

By Jessica Hagen | December 07, 2023 | 06:20 pm

SHARE f in t

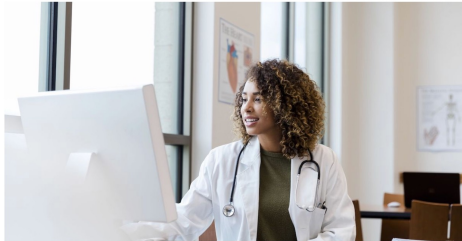


Photo: STEVE DEBENPORT/Getty Images

A study done by Google Research in collaboration with Google DeepMind reveals the tech giant developed an LLM with conversational and collaborative capabilities that can provide an accurate differential diagnosis (DDx) and help improve clinicians' diagnostic reasoning and accuracy in diagnosing complex medical conditions.

[17%] 244.9 Unspecified hypothyroidism [43%] one touch "ultra" test

Read aalim.kaiser.org



# What are the clinical needs of payers/back office/CIO/IT?

- Pre-authorization
- Revenue cycle management
- Risk identification
  - 30-day Readmission prediction
- Auditing and compliance
- EMR/Healthcare informatics
- Peer review
- Clinical trial search
- Foundational Models
  - ✓ Statistical ML models
  - ✓ DNN
  - LM
  - LLM
  - VLM
  - ✓ LLM for Search
- ✓ FM Workflow applications
- ✓ Data science tools

# Peer Review - Aortic Stenosis study

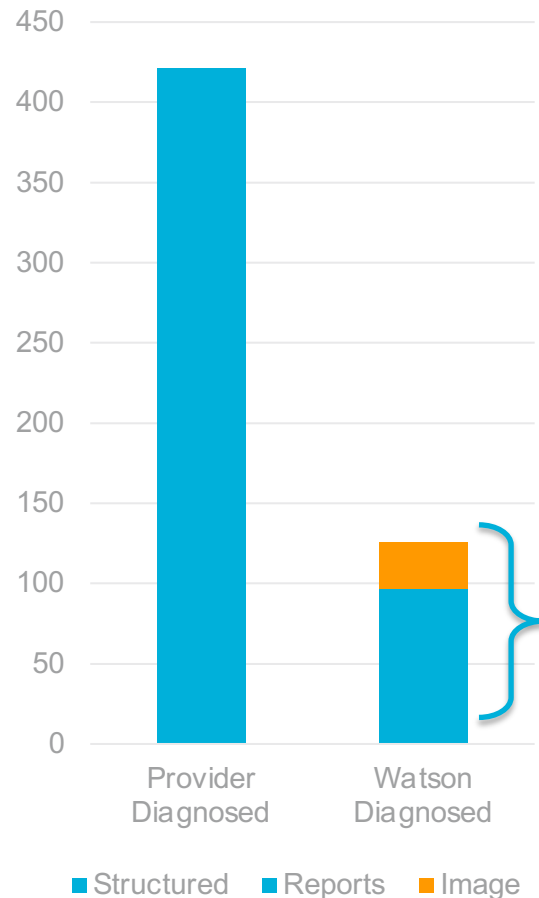
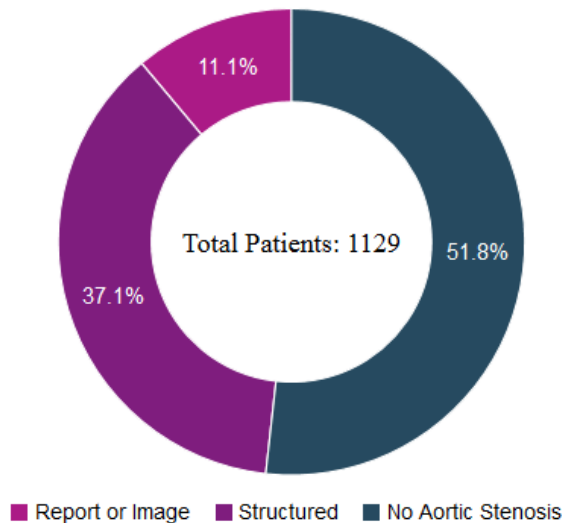
## Study 1 – at Kaiser hospitals

- Total Patients: 1,129
- Patients with AS: 547
- Diagnosed by Provider 421
- Undiagnosed or Tracked: 126
  - Reports: 97
  - Images: 29

## Study 2: At Sentara Hospitals

- Total Patients: 3090
- Patients with AS: 664
- Diagnosed by Provider 581
- Discrepancies found: 83

**Nearly 30% discrepancies found by ML models!**



# Clinical needs of FM for researchers

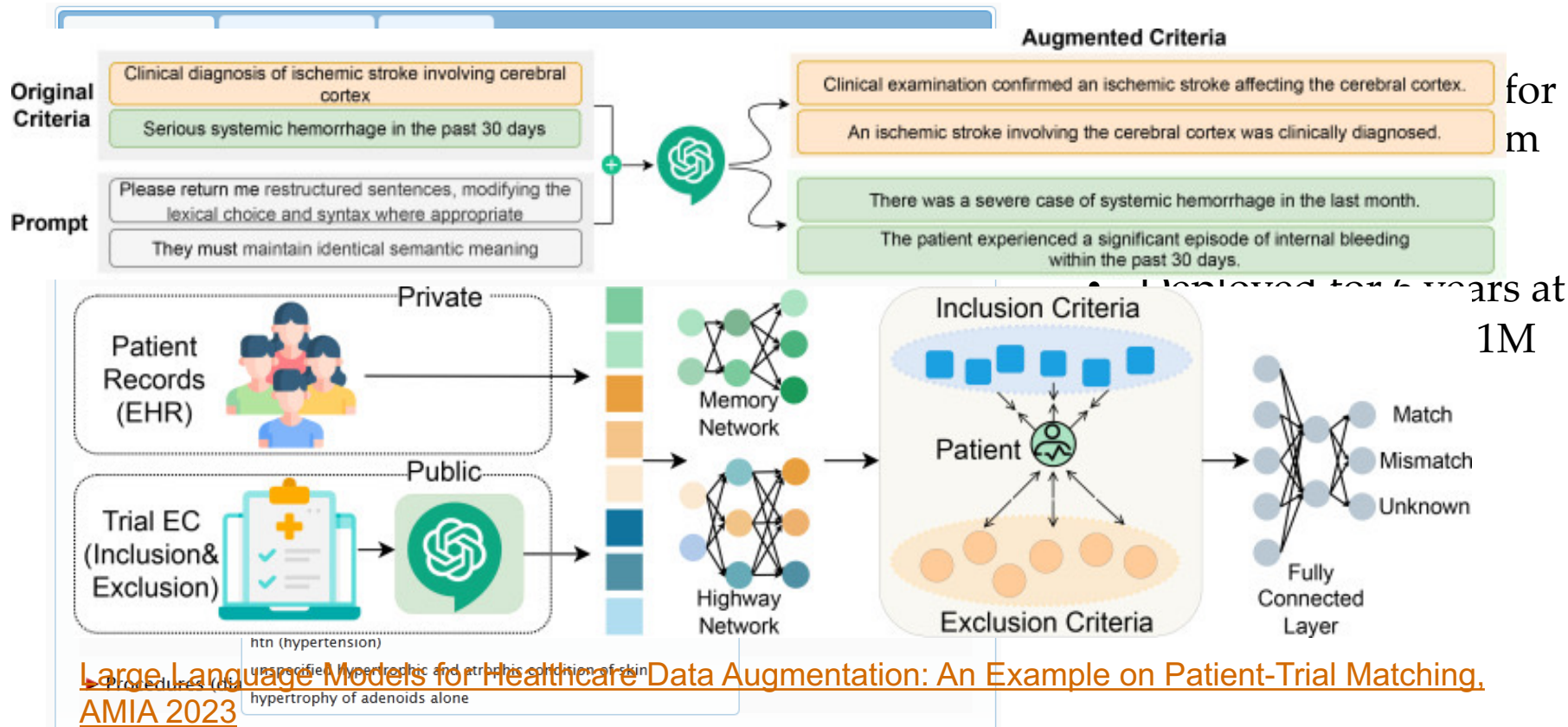
- Clinical Research and Education

- Compare effectiveness of treatments
- Predictive analytics
- Developing new guidelines for treatment
- Enhance medical curriculums
- Clinical trials research
- Accelerated discovery research (major area)

- Foundational Models

- ✓ Statistical ML models
- ✓ DNN
- ✓ LM
- ✓ LLM
- ✓ VLM
- ✓ FM for search
- ✓ Multimodal fusion models
- ✓ FM Workflow applications
- ✓ Data science tools

# Searching patients for clinical trials



# Clinical needs for patients

- Remote monitoring alerts
- Mobile tracking
- Triage and referral routing
- Patients-like-me
- Foundational Models
  - ✓ Statistical ML models
    - DNN
    - LM
    - LLM
    - VLM
    - FM for search
  - FM Workflow applications
    - ✓ Data science tools
    - ✓ Anomaly detection

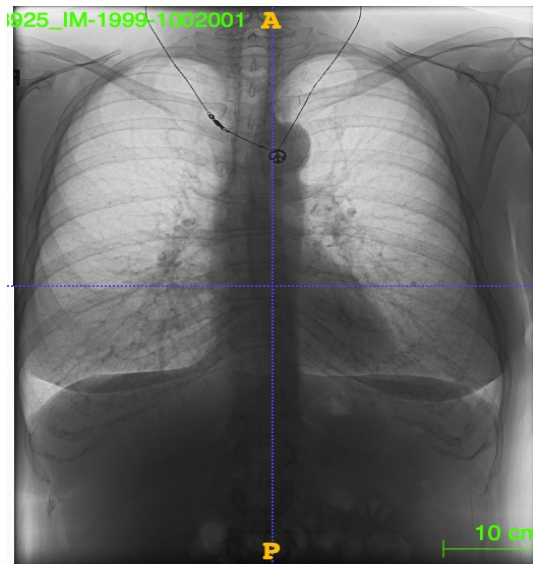
# Most relevant use case of FM

- Automated report generation
  - From dictation to report
  - From image to report
- How good are the current automated report generation using LLM?
  - XrayGPT, GPT4,..
  - Hallucinations
- A case study of automated reporting
  - Turing study on chest X-ray preliminary reads (preview)



# A Turing Test for Radiology AI

Chest X-ray



Automated Preliminary Read

There are **atherosclerotic** changes of the aorta.  
There are **calcified right** hilar and mediastinal lymph nodes.  
**Arthritic** changes of the skeletal structures are noted.