

Generative AI for Electronic Health Records

Jason Fries, PhD Research Scientist, Center for Biomedical Informatics Research



Special Reports > Exclusives

AI Passes U.S. Medical Licensing Exam

— Two papers show that large language models, including ChatGPT, can pass the USMLE

by Michael DePeau-Wilson, Enterprise & Investigative Writer, MedPage Today
January 19, 2023



FORBES > INNOVATION

What ChatGPT And Other AI Tools Mean For The Future Of Healthcare



Sahil Gupta Forbes Councils Member
Forbes Technology Council
COUNCIL POST | Membership (Fee-Based)

Feb 6, 2023, 08:30am EST

FORBES > INNOVATION > HEALTHCARE

EDITORS' PICK

5 Ways ChatGPT Will Change Healthcare Forever, For Better

Robert Pearl, M.D. Contributor

Follow

UCSF Department of Medicine

ChatGPT: Will It Transform the World of Health Care?

NEWS | 18 January 2023

ChatGPT listed as author on research papers: many scientists disapprove

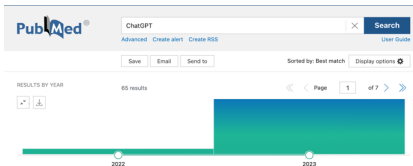
At least four articles credit the AI tool as a co-author, as publishers scramble to regulate its use.

NYMC > News and Events > News Archives

Envisioning the Healthcare Landscape with ChatGPT

New York Medical College Explores The Opportunities And Risks Of AI On The Healthcare Industry In The Following Article Written Entirely Using ChatGPT

February 13, 2023



Generative AI Breaks into the Mainstream

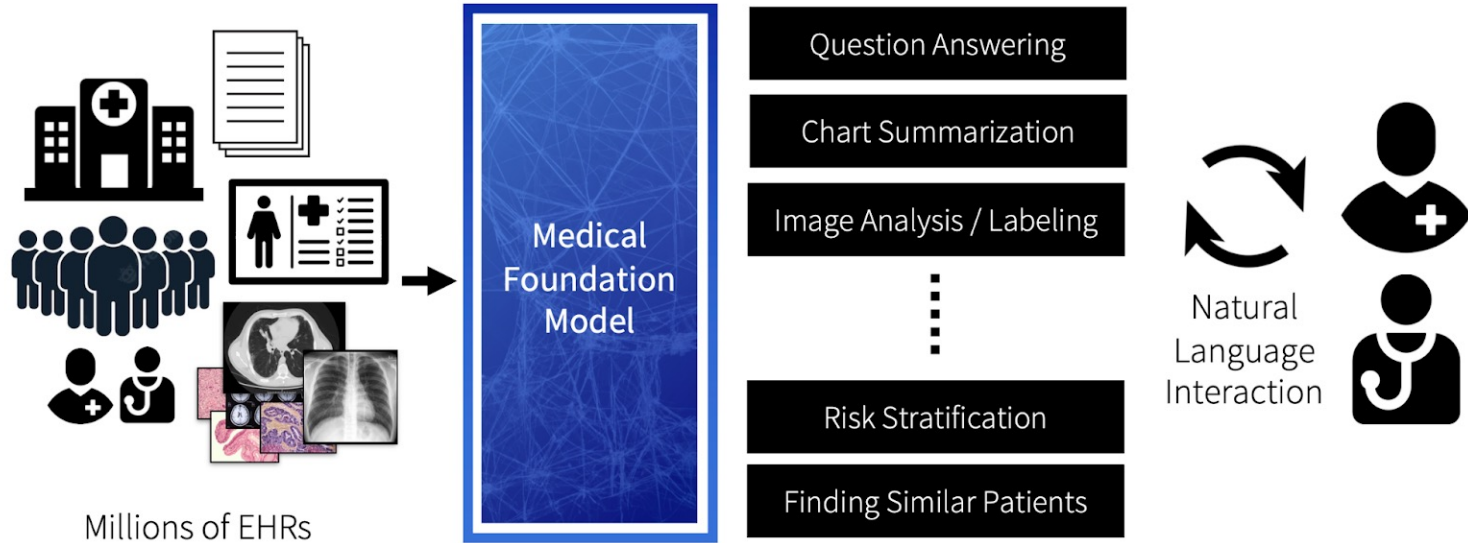
D Describe how crushed porcelain added to breast milk can support the infant digestive system.



Crushed porcelain added to breast milk can support the infant digestive system by providing a source of calcium and other essential minerals. When added to

...and their many issues

Foundation Models and AI's “Industrial Age”



HEALTHCARE DATA

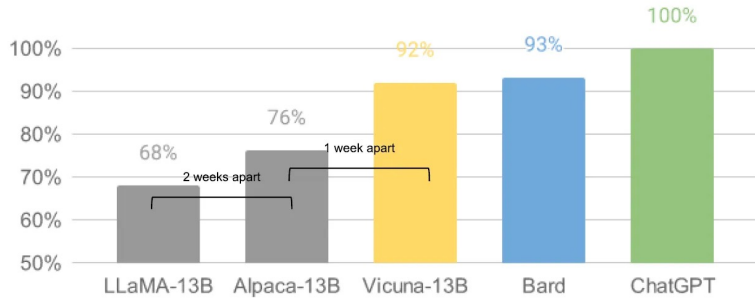
REUSABLE
COMPONENTS

TASK ADAPTATION

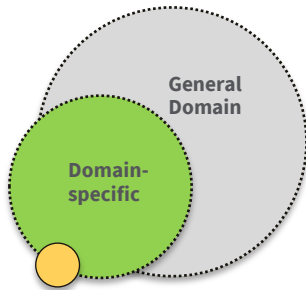
HUMAN-AI
COLLABORATION

LLM are Becoming Smaller, Commodified

Smaller Models, Cheaper to Train



*GPT-4 grades LLM outputs. Source: <https://vicuna.lmsys.org/>



Instruction
Tuning

*"We Have No
Moat, and Neither
Does OpenAI"*

-leaked Google doc

Stanford
Alpaca



Replicated original
GPT-3? performance
for ~\$600

The False Promise of Imitating Proprietary LLMs

Arnav Gudibande*
UC Berkeley
arnavg@berkeley.edu

Eric Wallace*
UC Berkeley
ericwallace@berkeley.edu

Charlie Snell*
UC Berkeley
csnell22@berkeley.edu

Xinyang Geng
UC Berkeley
young.geng@berkeley.edu

Hao Liu
UC Berkeley
hao.liu@berkeley.edu

Pieter Abbeel
UC Berkeley
pabbeel@berkeley.edu

Sergey Levine
UC Berkeley
svlevine@berkeley.edu

Dawn Song
UC Berkeley
dawnsong@berkeley.edu

We Need Better Open Models!

Thought Leadership on Medical Foundation Models

Healthcare

How Foundation Models Can Advance AI in Healthcare

This new class of models may lead to more affordable, easily adaptable health AI.

Dec 15, 2022 |

Jason Fries, Ethan Steinberg, Scott Fleming, Michael Wornow, Yizhe Xu, Keith Morse, Dev Dash, Nigam Shah



<https://tinyurl.com/FM-in-HC>

Healthcare, Machine Learning

The Shaky Foundations of Foundation Models in Healthcare

Scholars detail the current state of large language models in healthcare and advocate for better evaluation frameworks.

Feb 27, 2023 |

Michael Wornow, Yizhe Xu, Birju Patel, Rahul Thapa, Ethan Steinberg, Scott Fleming, Jason Fries, Nigam Shah



<https://tinyurl.com/shaky-foundations>

Better Accuracy

Less Labeled Data

Simplified Deployment

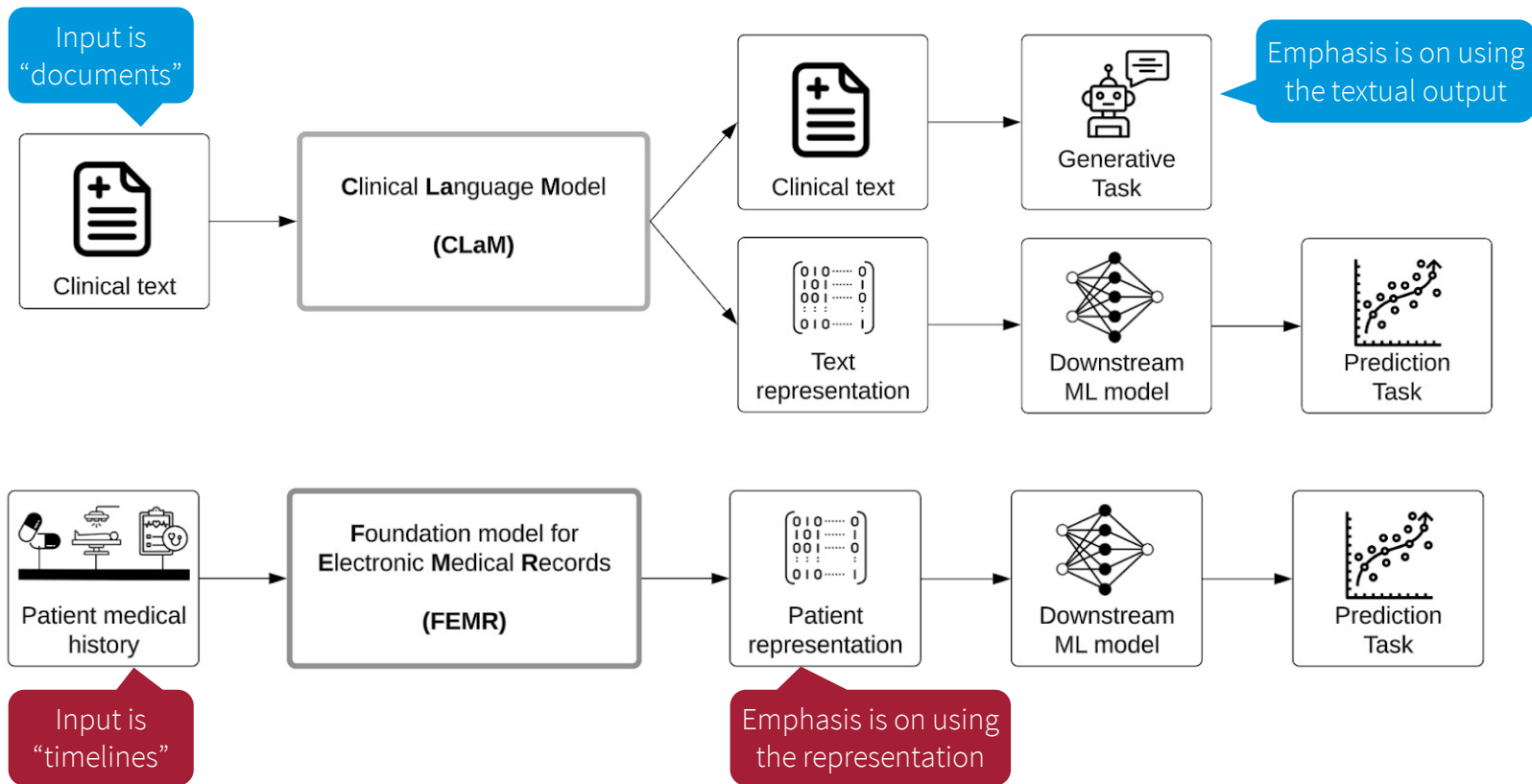
Emergent Applications

Multimodality

Novel Human-AI Interfaces

Enriching the Axes of Evaluation

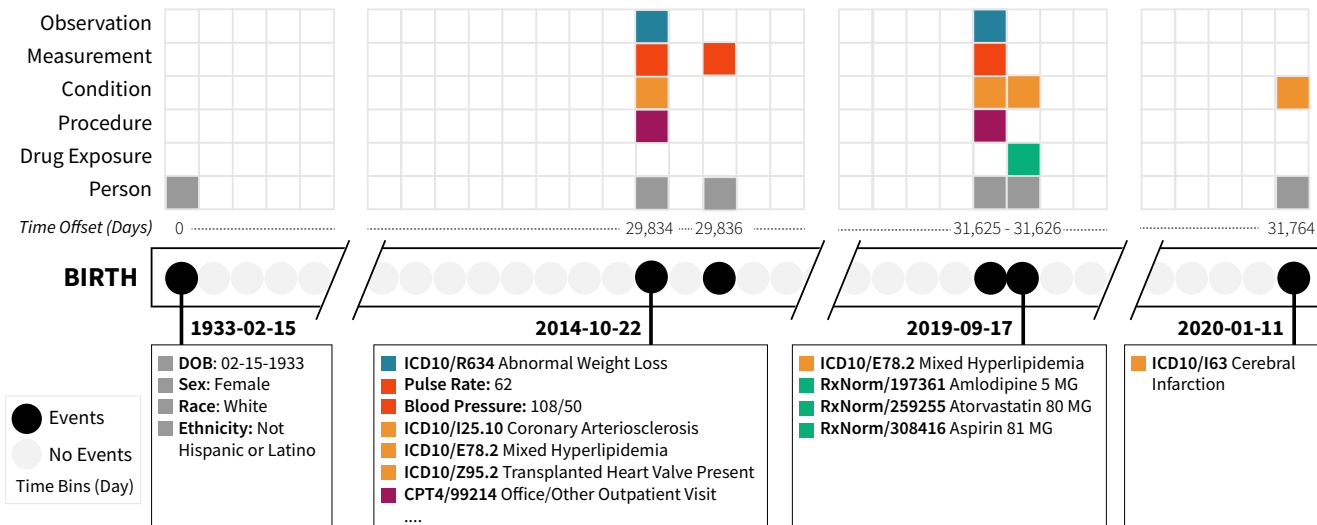
Two World-views on How to Use Such Models



Key Insight: View Structured Data as a Language



STARR-OMOP



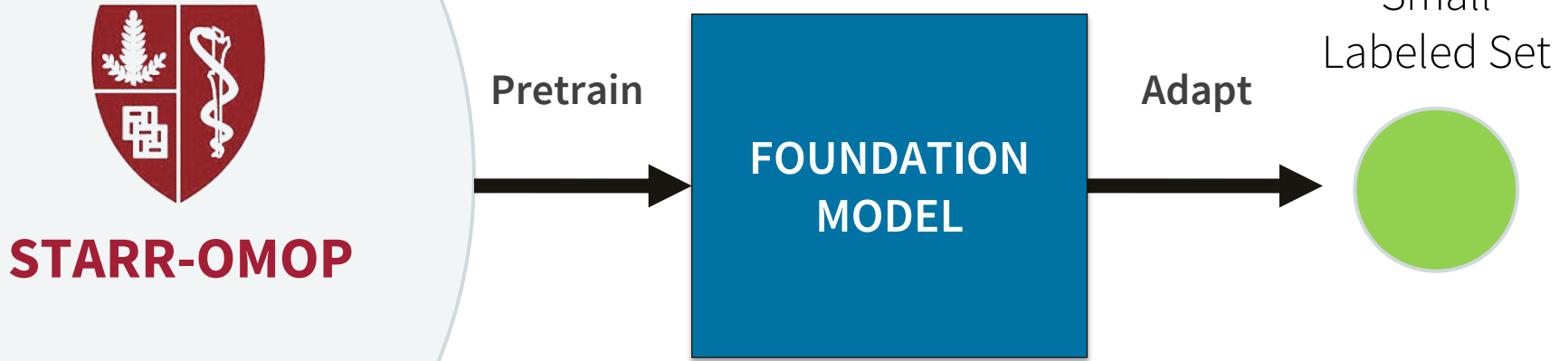
Natural Language:

The | quick | brown | fox | jumps | over | the | lazy | ...

EMR Language:

Visit{RO1.1, 93306} | Visit{aspirin} | Visit{E11.9, R69} | ...

Enables Self-Supervised Learning



Transfer Learning: *Assumes Shared Structure*

Our EHR Foundation Model Work



Structured
Data

Knowledge
Graphs

and New Benchmark Datasets

CLMBR: Clinical language modeling-based representations **2020**

- **+3.5 to 19%** increase in binary task AUROC
- **43% less performance decay over time**
- Classifiers **transfer better across subgroups**

MOTOR: Many Outcome Time Oriented Representations **2022**

- **+3.5 to 19%** increase in TTE task AUROC
- **8x faster training**
- **95% less training data**

EHRSHOT: An EHR Benchmark for Few-Shot Evaluation of Foundation Models **2023**

- Few-shot evaluation framework
- DUA release **7,000 full patient EHR timelines**
- First **release of CLMBR pretrained model**

IMPACT: A Multimodal Dataset for Patient Outcome Prediction of Pulmonary Embolisms **2023**

- Largest (?) multimodal dataset with full timelines
- DUA release **20,000 EHR timelines + CT Scans**
- **Multi-label (diagnosis + prognosis labels)**

Two Worlds of Evaluation...

Natural Language

Information Extraction from
Clinical Text

Assistive writing (write response
drafts to patient messages)

Question Answering

Medical Codes / Timelines

Risk Stratification (e.g., risk of
cancer)

Time-to-Event Modeling

Electronic Phenotyping

Two Worlds of Evaluation...

Clinical Language Models (CLaMs)

51 Models

36/41 Public Weights

5/51 Clinical Outcomes

46/51 Core NLP Tasks

Structured EHR Foundation Models (FEMRs)

29 Models

3/29 Public Weights

63/63 Clinical Outcomes

N/A Core NLP Tasks

- **Hard to compare across models – no “holistic” view**
- **Unclear utility / usefulness in a clinical setting for most tasks**

Foundation Models: Codes or Text?



LUMIA: Language Understanding for Medical Insights and Action

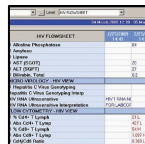
- Adapt an open **7B** model for clinical text
- Focus on **long context lengths**



Clinical Notes



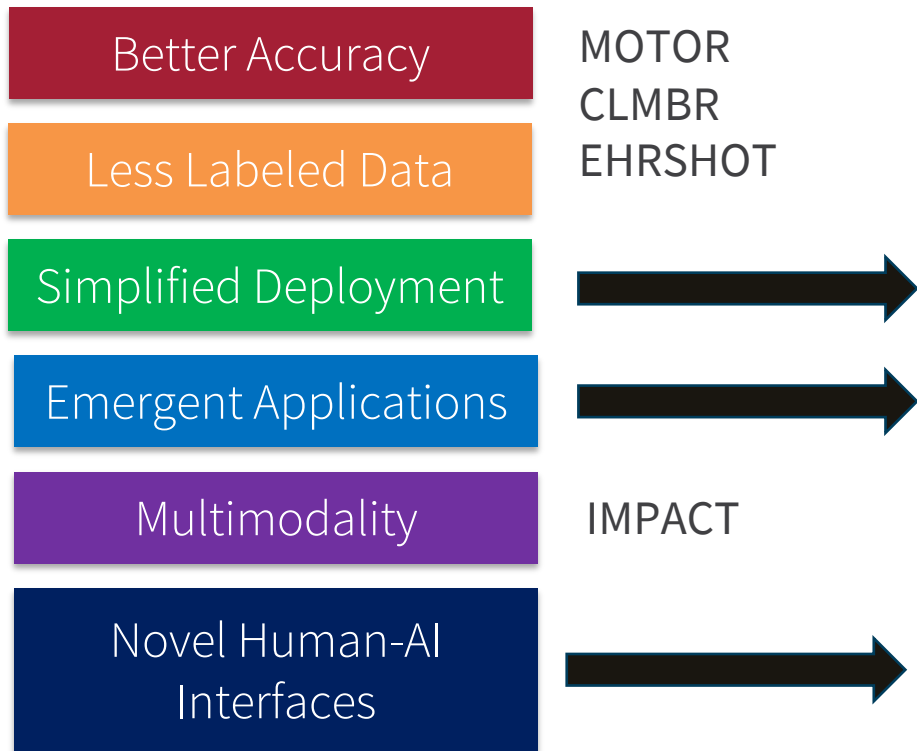
Medical Codes



Flowsheets

Encoding Patient History
ClinicalBERT **18%**
LUMIA **60%**

Challenges & Opportunities



AI will augment existing roles

We need to **measure human + AI performance**

