NLP for Health

CSCI 4907/6907

Aya Zirikly
08/27/2025

Logistics

Time

- Every Wednesday 12:45PM - 03:15PM
- 08/25/2025 - 12/08/2025 (excluding Wednesday 11/26/2025 for Thanksgiving break)
- Last class 12/03/2025

Location

- ROME Hall (2300 I Street NW), B103

Website

- https://azirikly.github.io/nlp4health-fall25/

Logistics

Grading

- Paper presentations 30%
- Participation 20%
- Class project 50%

Office hours

- TBD

University policies

- Student affairs https://students.gwu.edu/
- Academic Integrity https://students.gwu.edu/code-academic-integrity

Paper presentations sign-up

Please make sure you sign-up for ~2 paper presentations throughout the semester

https://docs.google.com/spreadsheets/d/1lEqBHbOeBPB_fqHuqNmTzBMYt8EY7wGAZYsaO-7H8Lc/edit?usp=sharing

Project

- ~3 persons per team
- Project
  - Proposal
  - Presentation
  - Documentation and code (github repo)
  - Extra credit for paper publication

Project

Pipeline

- Dataset
- Identify tasks and research questions
- Preprocessing
- Modeling
- Evaluation
- Analysis
- *Publication*

Medical data

- ~1.0 billion physician office visits annually
  - ~50.3% of these visits are primary care physicians
- 88.2% of office-based physicians in the U.S. use some form of EMR/EHR system, and 77.8% use certified systems (based on CDC)
- May 2020-April 2023, outpatient providers in Epic EHR systems alone generated 1.7 billion clinical notes
  - These notes were written by 166,318 outpatient providers
  - This suggests that well over three-quarters of patient visits are likely to include structured documentation, including text notes.

What is EHR

An Electronic Health Record (EHR)* is an electronic version of a patient's medical history, maintained by the provider over time.

- Demographics
- Progress notes
- Problems
- Medications
- Vital signs
- Past medical history
- Immunizations
- Lab data
- Radiology reports

* Centers for Medicare & Medicaid Services (CMS)

Outpatient vs. inpatient

Outpatient visit

- Medical services provided without an overnight stay
  - Routine checkups

Inpatient visit

- Care provided when a patient is admitted to a hospital or healthcare facility for at least one overnight stay
  - Childbirth

Types of EHR

Basic EHR

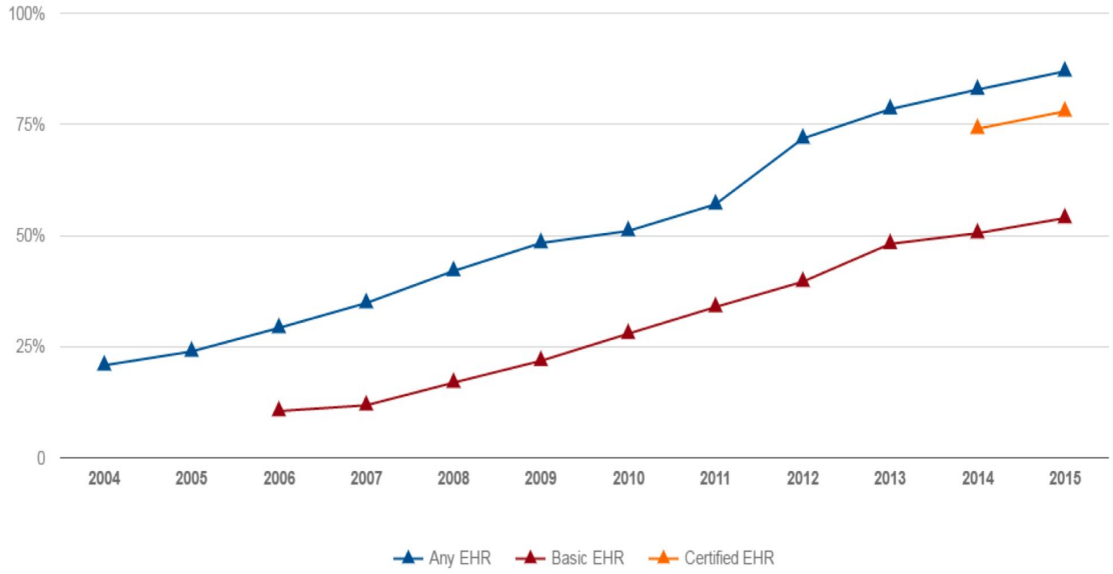
Offers fundamental patient data management

Certified EHR

Meets regulatory standards for enhanced functions

EHR

Comprehensive digital patient records system


100%
75%
50%
25%
0
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
Any EHR
Basic EHR
Certified EHR

Example

Patient Name: John Doe
Date: August 26, 2025
Provider: Dr. Jane Smith
Visit Type: Follow-up for hypertension

S – Subjective
Patient reports feeling generally well but notes occasional headaches in the morning. Denies chest pain, shortness of breath, or dizziness. Admits to missing doses of medication twice last week due to travel.

O – Objective

- BP: 148/92 mmHg
- HR: 78 bpm
- Temp: 98.6°F
- Weight: 198 lbs
- Labs from last visit: Creatinine 1.1, Potassium 4.2, HbA1c 6.8%

A – Assessment

- Hypertension, not well controlled
- Non-adherence to medication regimen likely contributing
- No signs of end-organ damage at this time

P – Plan

- Reinforce medication adherence
- Adjust lisinopril dose from 10 mg to 20 mg daily
- Schedule follow-up in 4 weeks
- Encourage home BP monitoring and logging
- Discuss lifestyle modifications (diet, exercise)

Platforms

Managing health records

- Epic
- Oracle Health (Cerner)

Functionalities

- Templates
- Dropdown menus
- Fill in functionalities
- AI features

Why clinical notes are important

- Improve patient care
- Find critical information from past visits
- Suggest diagnosis or medications
- Ensure compliance and follow-ups

If we have access to thousands / millions of medical records representing millions of patients what can we learn?

Advances

- Make research advances in medicine
- Discover negative medication interactions
- Identify a cohort of patients for a research study
- Ensure patient safety procedures

Clinical text

- Most EHR data is structured
  - Relatively easy to export and use
- Clinical text is unstructured?
  - How do we use it?
  - Natural Language Processing

Challenges with clinical text

- Access
  - HIPAA (Health Insurance Portability and Accountability Act) compliance
- Domain specific language
  - Medical jargon (abbreviations, shorthand)
- Noisy text
  - Grammar issues, templates, fragmented sentences, copy-paste issues
- Ambiguity and context dependence
  - Cold: symptom vs. temperature
- Data sparsity and imbalance
  - Rare conditions and treatments
- Generalization

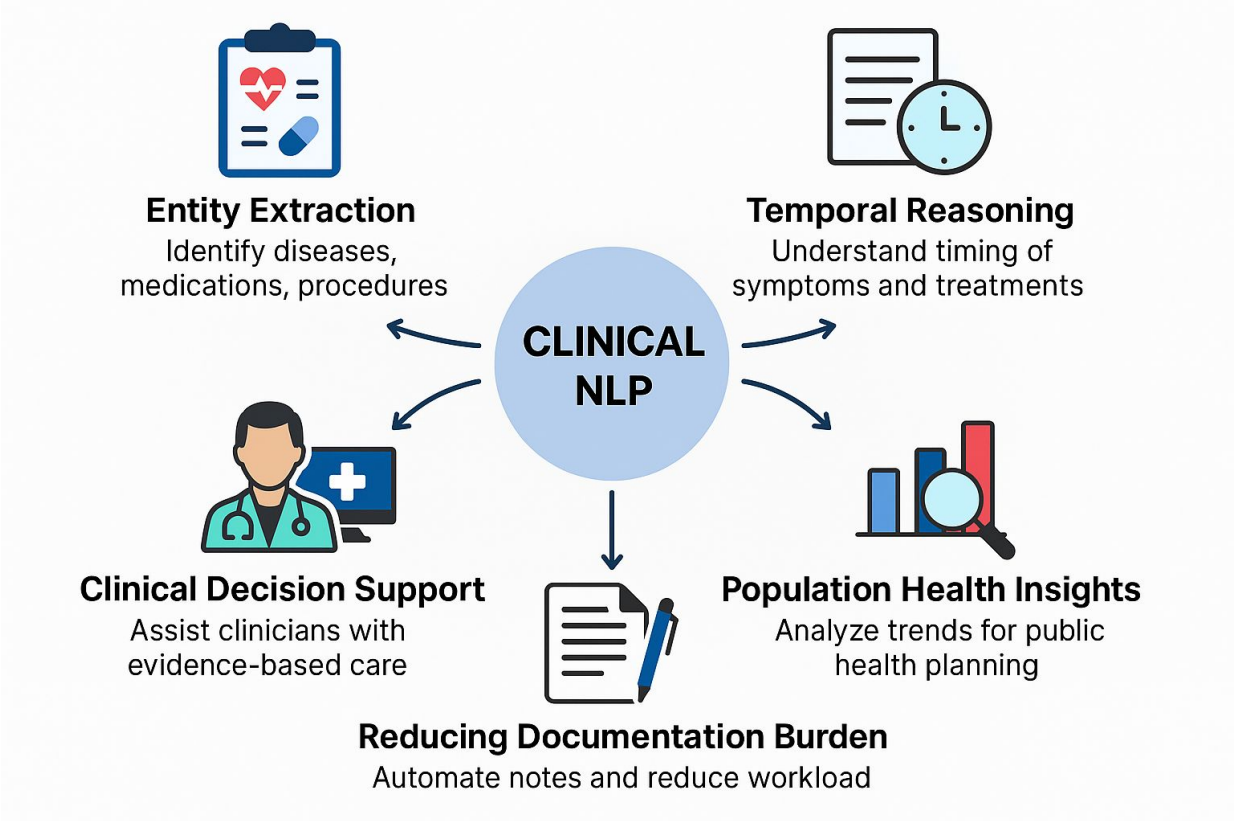
Clinical text alternatives?

Clinical text alternatives

- Social media
  - Easier access
  - Stand-alone or supplementary to clinical data
  - Broader population insights
  - Lack of outcome
  - Lower documentation bias
- Synthetic data
  - Data augmentation
  - Benchmarking
  - Imbalanced dataset

Clinical Natural Language Processing

- Applying NLP tools to clinical text


Entity Extraction
Identify diseases, medications, procedures
Temporal Reasoning
Understand timing of symptoms and treatments
CLINICAL NLP
Clinical Decision Support
Assist clinicians with evidence-based care
Population Health Insights
Analyze trends for public health planning
Reducing Documentation Burden
Automate notes and reduce workload

Entity and information extraction

Genomic Research & Precision Medicine

- Extracting genetic and phenotypic entities from clinical text to support precision medicine.
- Entities Extracted: Genes, mutations, disorders, symptoms.
- Improves clinical decision-making by linking patient data to genomic insights.

Entity and information extraction

Clinical Decision Support

- Extracting structured data from unstructured notes to assist in treatment recommendations.
- Entities Extracted: Diagnoses, medications, lab results, temporal markers.
- Supports timely and evidence-based clinical decisions.

Approaches

- Traditional approaches
- SVM, logistic regression
- Deep learning
- CNN, forward neural networks, ..
- Large Language Models

Clinical NLP tools

- cTAKES (clinical Text Analysis and Knowledge Extraction System)
  - Mayo Clinic.
  - Extracts information like diseases, medications, procedures, and anatomical sites from clinical notes.
  - Uses dictionary-based and machine learning methods.
- MetaMap
  - National Library of Medicine (NLM).
  - Maps biomedical text to UMLS Metathesaurus concepts.
  - Concept normalization and semantic analysis.

Clinical NLP tools

- CLAMP (Clinical Language Annotation, Modeling, and Processing)
  - UTHealth.
  - Supports named entity recognition, assertion detection, and relation extraction.
- NegEx
  - A simple rule-based algorithm for detecting negation in clinical text.
  - Often integrated into other NLP pipelines like cTAKES and MetaMap.

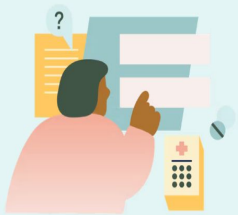
LLM for health

- Clinical Decision Support
- EHR enhancement
  - Free to structured text
- Automated patient communication
- Predictive health outcomes
- Personalized treatment plans
- Training and education

Buoy Health (diagnosis)

1

Use our A.I. to figure out what to do



Your situation is specific to you. Our A.I. asks questions and helps narrow down what is going on based on your answers. No more guessing and filling in the blanks. It's like talking to your friend who happens to be a doctor.

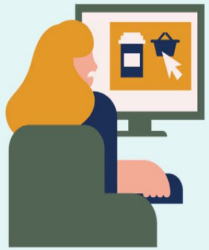
2

Read our content written by clinicians

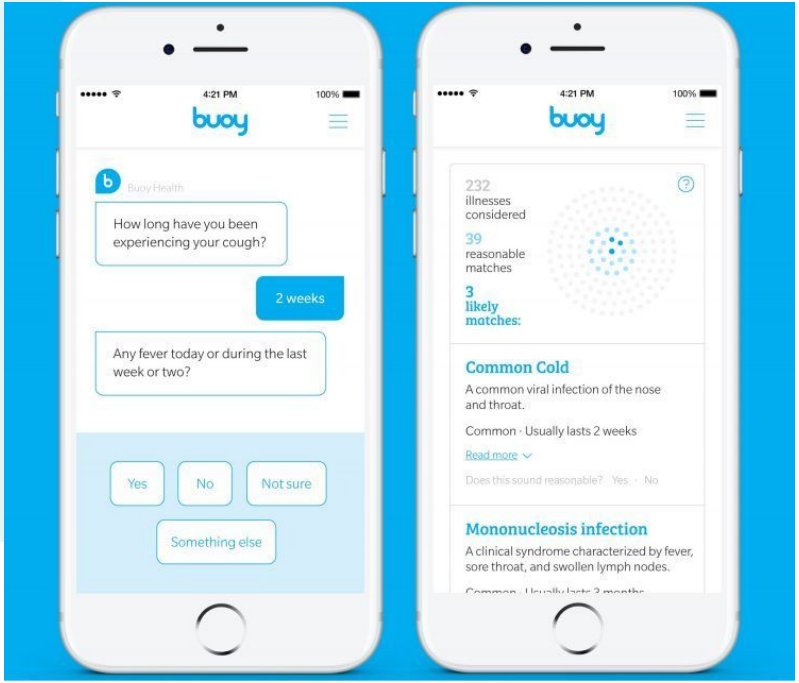
Wouldn't that be great if the content you read was actually scientifically vetted by real experts? Look no further.

3

Check out our reviews and Top 10 lists



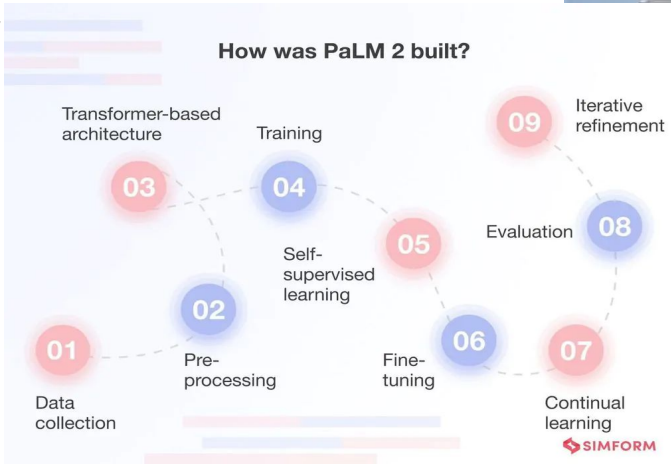
Now that we've helped you narrow things down with both A.I. and medical experts, find the right product/service/doctor for you with the help of our team of reviewers.



https://www.buoyhealth.com/

Med-PaLM

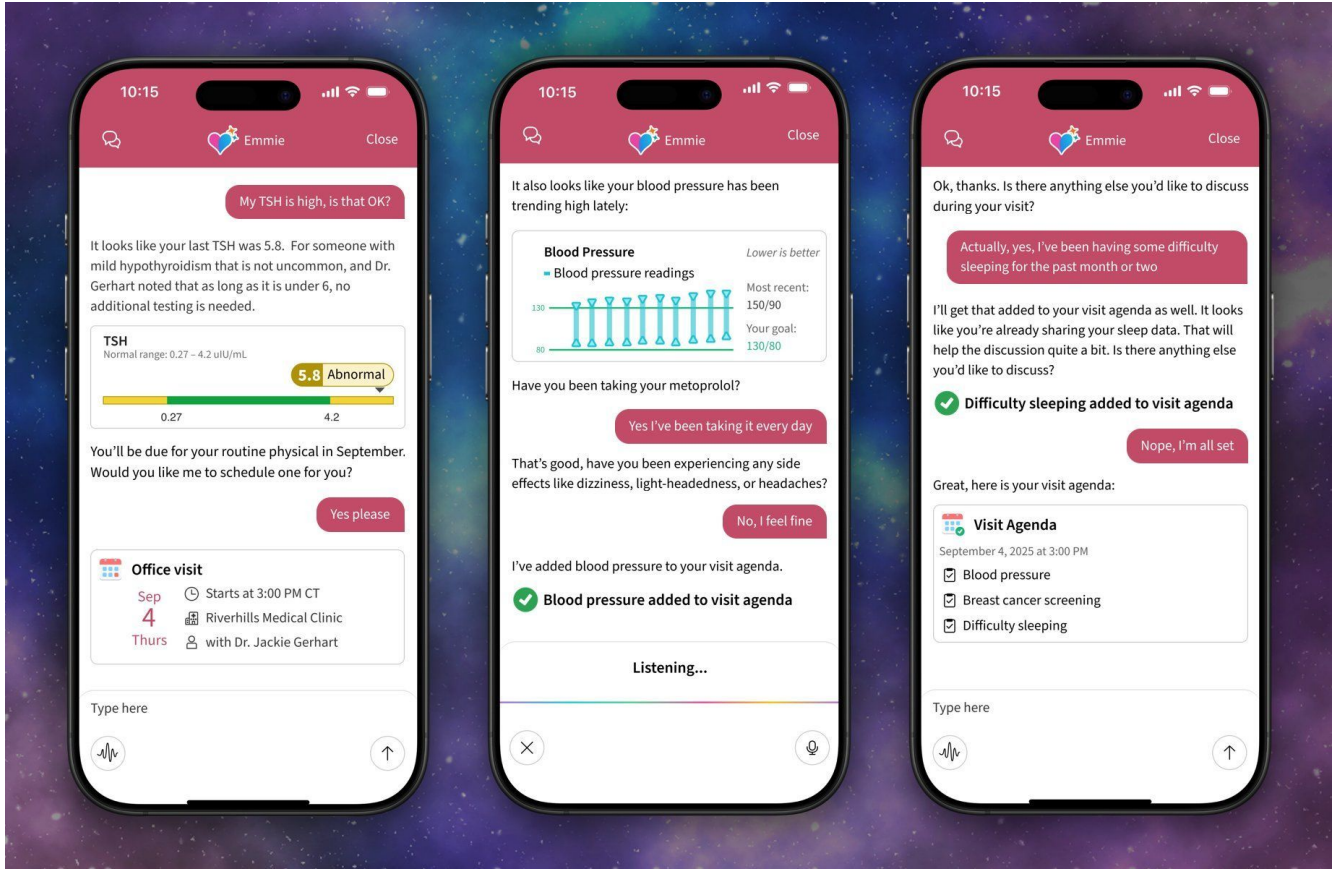
- Google
- Question answering system for medical domain
- 86.5% accuracy



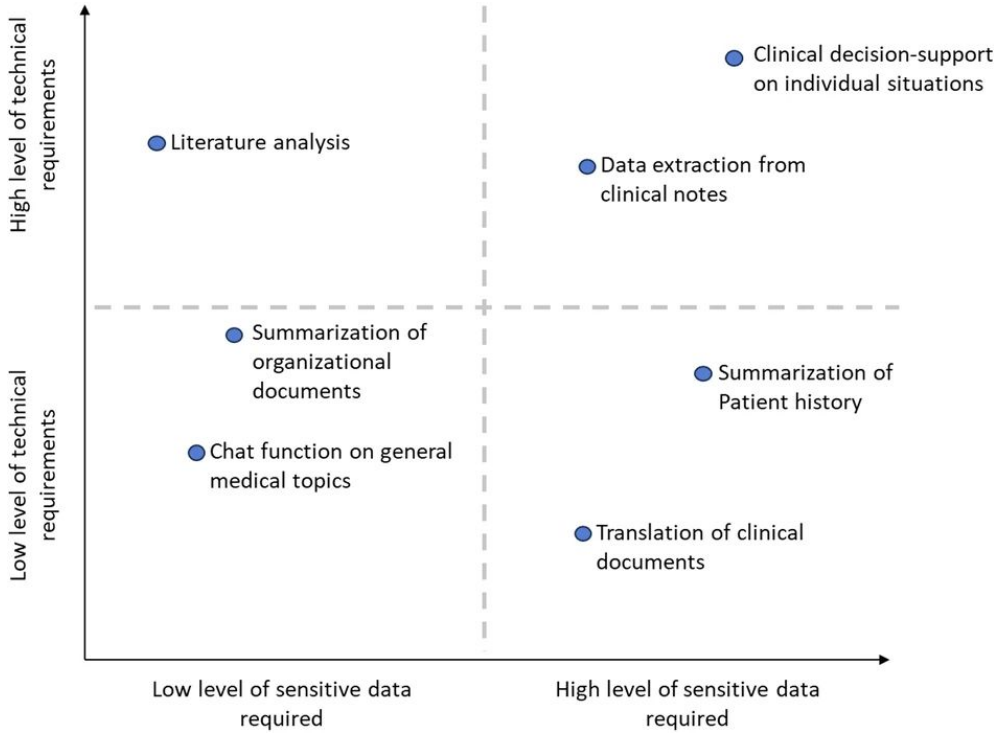
Lauren Parr, 2023, Revolutionizing Healthcare: Google's New AI Model Med-PaLM 2
Hiren Dadhuk 2023, Exploring the Potent Magic of PaLM 2: A Definitive Guide

Emmie

- Integrated with Epic
- August 2025
- virtual assistant designed to help patients manage their health care.
- Support patient
  - Explaining test results in easy-to-understand terms
  - Suggesting next steps, or guiding patients through open-ended conversations about their health
- Support clinicians
  - Pre-visit summaries
  - Taking real-time notes
  - Placing orders or verifying prior authorization requirements


10:15
Emmie
Close
My TSH is high, is that OK?
It looks like your last TSH was 5.8. For someone with mild hypothyroidism that is not uncommon, and Dr. Gerhart noted that as long as it is under 6, no additional testing is needed.
TSH
Normal range: 0.27 – 4.2 uIU/mL
5.8 Abnormal
0.27
4.2
You'll be due for your routine physical in September. Would you like me to schedule one for you?
Yes please
Office visit
Sep
4
Thurs
Starts at 3:00 PM CT
Riverhills Medical Clinic
with Dr. Jackie Gerhart
Type here
10:15
Emmie
Close
It also looks like your blood pressure has been trending high lately:
Blood Pressure
Lower is better
Blood pressure readings
Most recent:
150/90
Your goal:
130/80
130
80
Have you been taking your metoprolol?
Yes I've been taking it every day
That's good, have you been experiencing any side effects like dizziness, light-headedness, or headaches?
No, I feel fine
I've added blood pressure to your visit agenda.
Blood pressure added to visit agenda
Listening...
10:15
Emmie
Close
Ok, thanks. Is there anything else you'd like to discuss during your visit?
Actually, yes, I've been having some difficulty sleeping for the past month or two
I'll get that added to your visit agenda as well. It looks like you're already sharing your sleep data. That will help the discussion quite a bit. Is there anything else you'd like to discuss?
Difficulty sleeping added to visit agenda
Nope, I'm all set
Great, here is your visit agenda:
Visit Agenda
September 4, 2025 at 3:00 PM
Blood pressure
Breast cancer screening
Difficulty sleeping
Type here

LLM for health



Dennstadt et al. 2025, Implementing large language models in healthcare while balancing control, collaboration, costs and security

Next week

- Medical Ontologies and UMLS
- Papers
  - TBD

Building familiarity

- Introduce yourself
- Major
- Degree
- Are you interested in research?
- What made you sign up for the class?
- Goals
- Topics of interest for the class project

BREAK 2:00- 2:15

Tutorial on Cerberus HPC

By: Jake Messick
Cyberinfrastructure Specialist
Research Technology Services