Large Language Models for Healthcare

Monica Agrawal

March 15, 2024

Disclosure / conflict of interest

I am a co-founder of Layer Health (financial interest)

Language is deeply embedded across medical workflows

Clinical Notes

Medical Literature

Trial Criteria

Communication + Instructions

…and connects disparate stakeholders.

Clinicians

Researchers

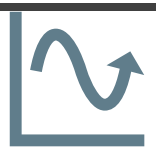
Patients

Potential of EHR data

Can learn from past data, stay up-to-date, and empower patients

- What drug would lead to the **best outcome** for **this patient**?

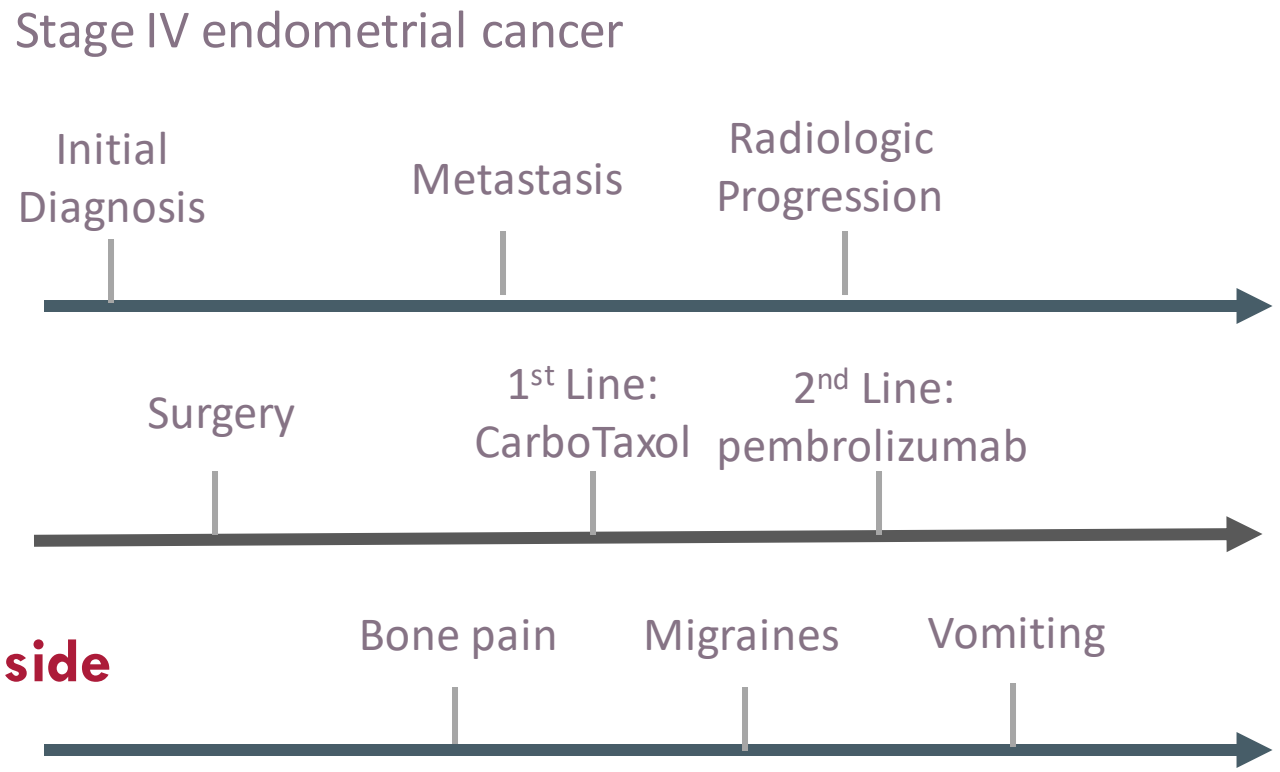
- Which **clinical trials** is my patient eligible for?



…but this data is often underutilized

Much of a patient's story is only in text

- **Disease**
- **Disease Status**
- **Interventions**
- **Symptoms/ side effects**
- **Confounders**



The information often doesn't exist elsewhere

Example: *Deviation* from medication plan

Pharmacy records say:

"The doctor ordered 6 weeks of medication starting February 1"

Notes say:

"The patient became nauseous and stopped taking the drug on February 5th"

Challenge: Difficulty of parsing clinical texts

…many ways to say the same thing

Fever

=

Febrile

=

Pyrexia

=

100+ degree temperature

Challenge: Difficulty of parsing clinical texts

…single way to say many things

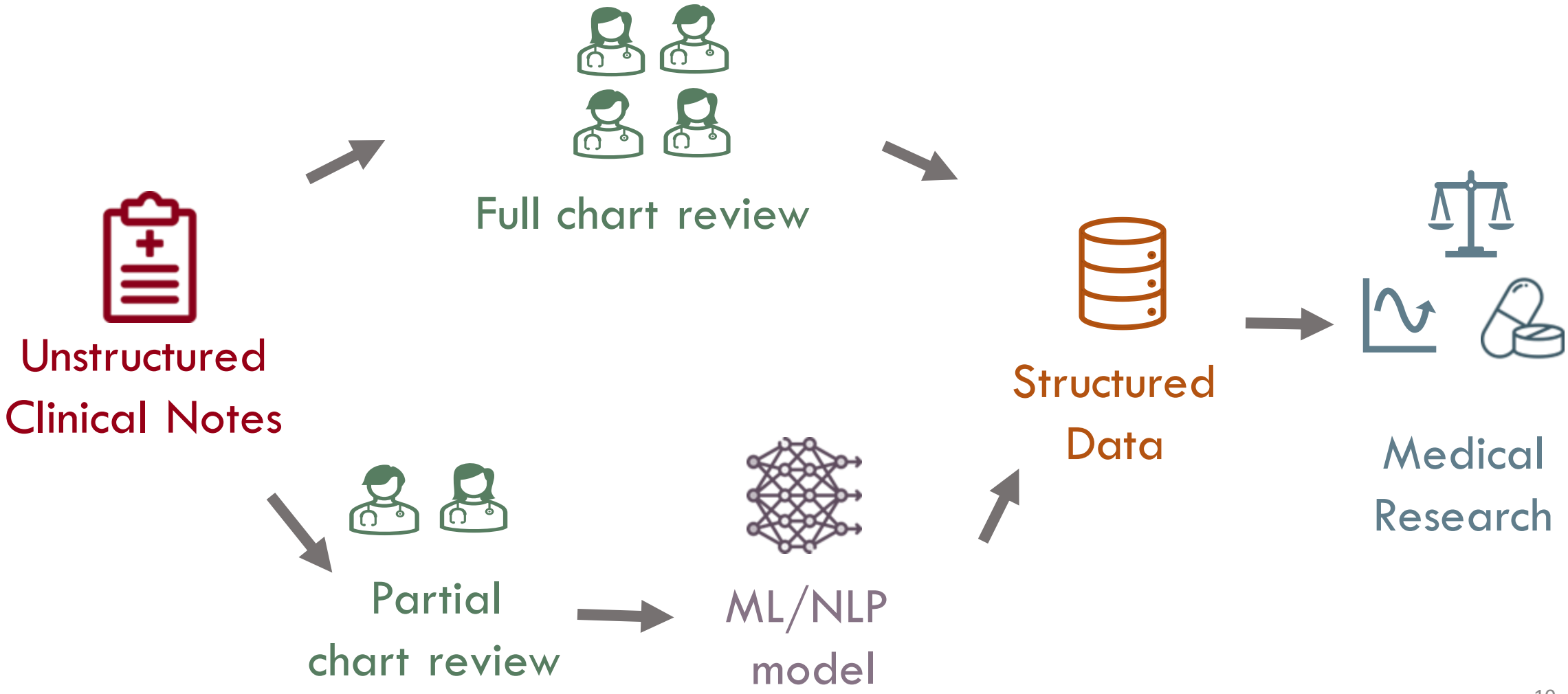
Challenge: Difficulty of parsing clinical texts

…traditionally hard for people and computers alike

Carboplatin / ~~Carbodome~~ — Intra-arterial / ~~Intra-articular~~ — Triple-negative breast cancer — discontinue / ~~D/C current~~ — ~~discharge~~

"On carbo ia for TNBC. Will dc."

Challenge #2: Status quo for information extraction



Opportunity: LLMs enable complex reasoning

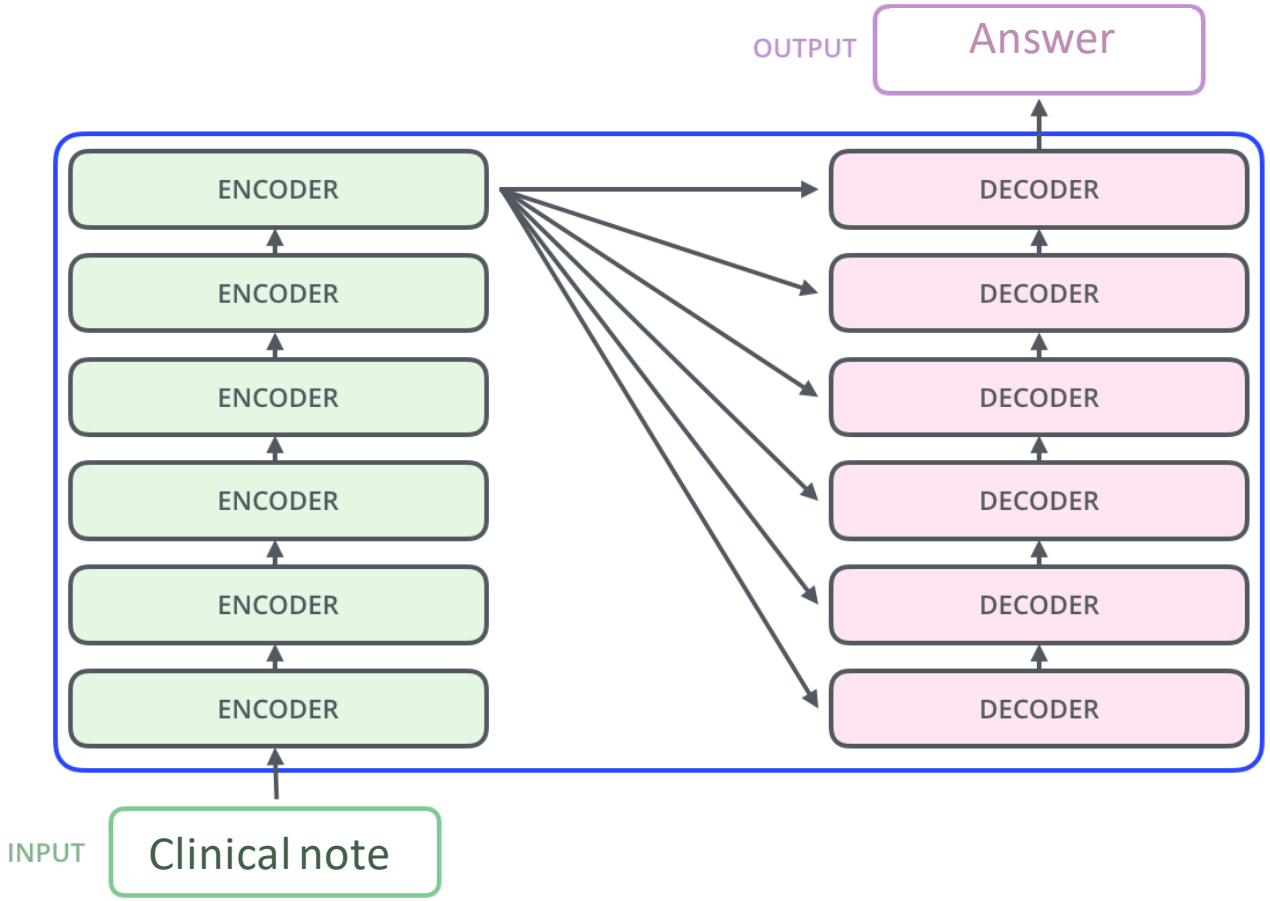
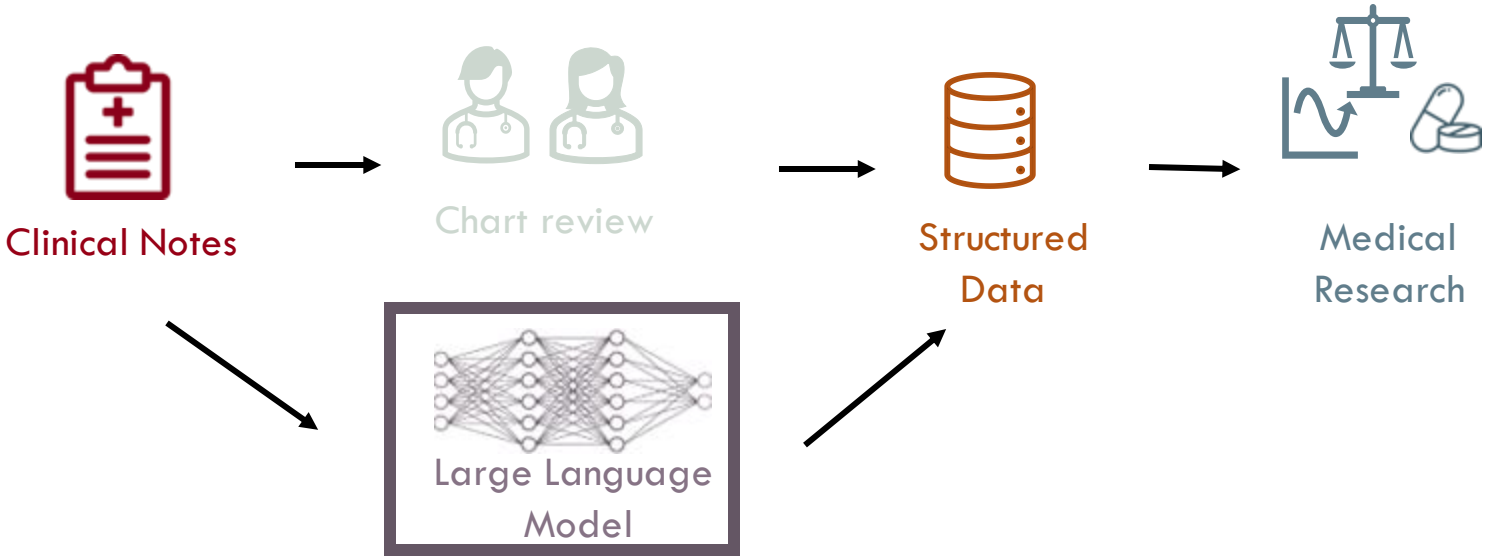


Image from Jay Alammar, Blog (2018). https://jalammar.github.io/illustrated-transformer/

Opportunity: Language models can help us parse text



Large Language Models are Few-Shot Clinical Information Extractors

*Empirical Methods in Natural Language Processing (**EMNLP**), 2022.*

Monica Agrawal, Stefan Hegselmann, Hunter Lang, Yoon Kim, David Sontag

Large Language Models are Few-Shot Clinical Information Extractors

1) Creation of clinical datasets for benchmarking LLMs

Large Language Models are Few-Shot Clinical Information Extractors

1) Creation of clinical datasets for benchmarking LLMs
2) Method for extracting complicated outputs from LLMs
   - Decreasing hallucinations
   - Improving performance of out-of-the-box models

Large Language Models are Few-Shot Clinical Information Extractors

1) Creation of clinical datasets for benchmarking LLMs
2) Method for extracting complicated outputs from LLMs
   - Decreasing hallucinations
   - Improving performance of out-of-the-box models
3) Results of LLMs on clinical information extraction
   - Clinical concept disambiguation
   - Clinical trial parsing
   - Clinical coreference resolution
   - Medication + status (active, discontinued) identification
   - Medication + attribute relation (dosage, reason) extraction

Example: Medication status identification

8. Albuterol 2 puffs every 4-6 hours as needed. HOSPITAL COURSE: This is an 80-year-old female who was hospitalized about 2 months ago for chronic obstructive pulmonary disease exacerbation…she was put on Augmentin ES and…

- medication: "Albuterol", dosage: "2 puffs", frequency: "every 4-6 hours", duration: "as needed"
-medication: "Augmentin ES", duration: "2 months"

Extraction of Social Determinants of Health

Guevara et al. examined the capabilities of LLMs for extracting SDoH

- Employment, housing, parent, relationship, social support, transport
- Evaluated both few-shot prompting of commercial models (GPT series) and fine-tuned models
- Their models identified 90%+ of patients with an adverse SDoH, compared to 2% by ICD codes

However, artificially injecting demographics (e.g. "person" → "Asian woman") did change answers ~10-20% of the time, indicating room for increased robustness

Guevara et al., Nature Digital Medicine 2024

How can we leverage LLMs towards equity?

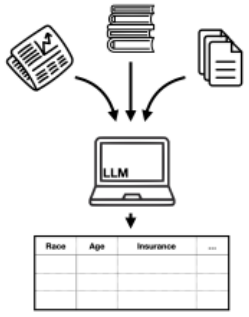
Use large language models to promote equity, *2023.*

Pierson, Shanmugam, Movva, Kleinberg, **Agrawal**, Dredze, Ferryman, Gichoya, Jurafsky, Koh, Levy, Mullainathan, Obermeyer, Suresh, Vafa

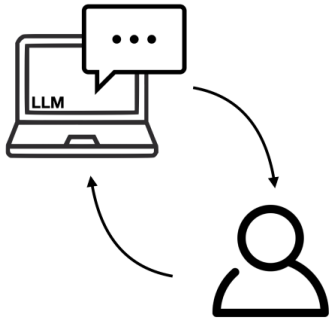
Opportunities for LLMs to promote equity

Detecting human biases 1

Creating structured datasets for equity research 2



Improving equity of access to information 3



Pierson, Shanmugam, Movva, Kleinberg *et al, 2023*

Improving access to information

- Access to clinical notes can enable:
  - Memory of appointment
  - Ability to further understand medical conditions
- Patients from **less privileged backgrounds** cite higher importance to this access
- Unfortunately **only 25% of patients** could adequately understand own surgical summary

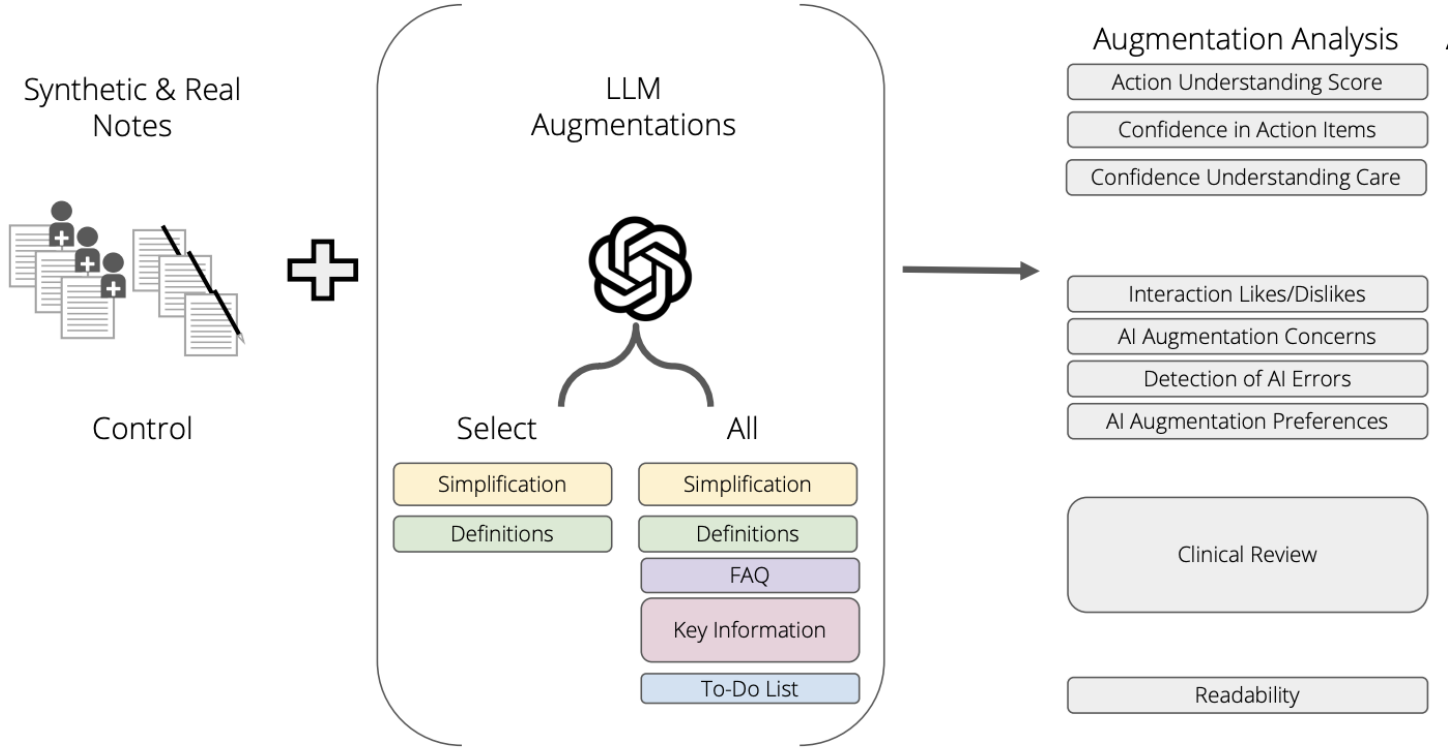
How can we make clinical notes more approachable for patients?

Gerard et al., JMIR 2018; Choudhry et al, Am J Surg 2016

Improving access to information

We investigated the impact of LLM assistance on comprehension of clinical notes

Understanding improved (including vulnerable subgroups), but errors persist

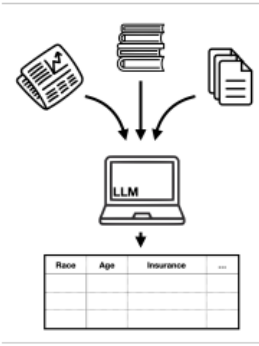


Mannhardt and Bondi-Kelly et al, 2024

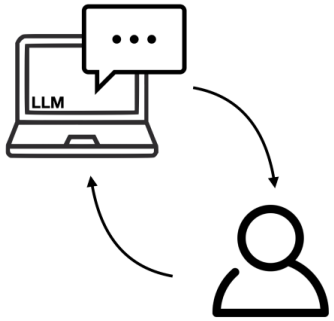
Opportunities for LLMs to promote equity

1 Detecting human biases

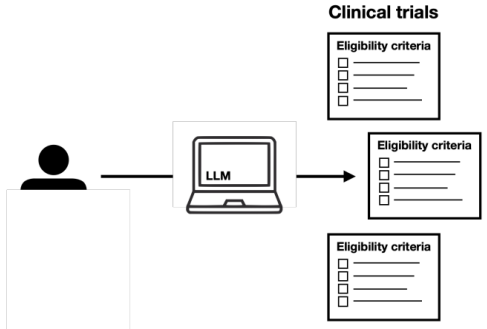
2 Creating structured datasets for equity research



3 Improving equity of access to information



4 Making matching systems more equitable



Pierson, Shanmugam, Movva, Kleinberg *et al, 2023*

Trial matching

AAMCNEWS

Clinical trials seek to fix their lack of racial mix

Most drugs have been tested primarily on White men, casting doubt about their efficacy for others. Researchers are trying to diversify who participates in studies.

By Patrick Boyle, Senior Staff Writer

Aug. 20, 2021

"Black patients account for just 5% of clinical trial participants in the US"

Boyle 2021, AAMC News; Alegria et al 2021, JAMA Network Open

Language is deeply embedded across medical workflows

Clinical Notes

Medical Literature

Trial Criteria

Communication + Instructions