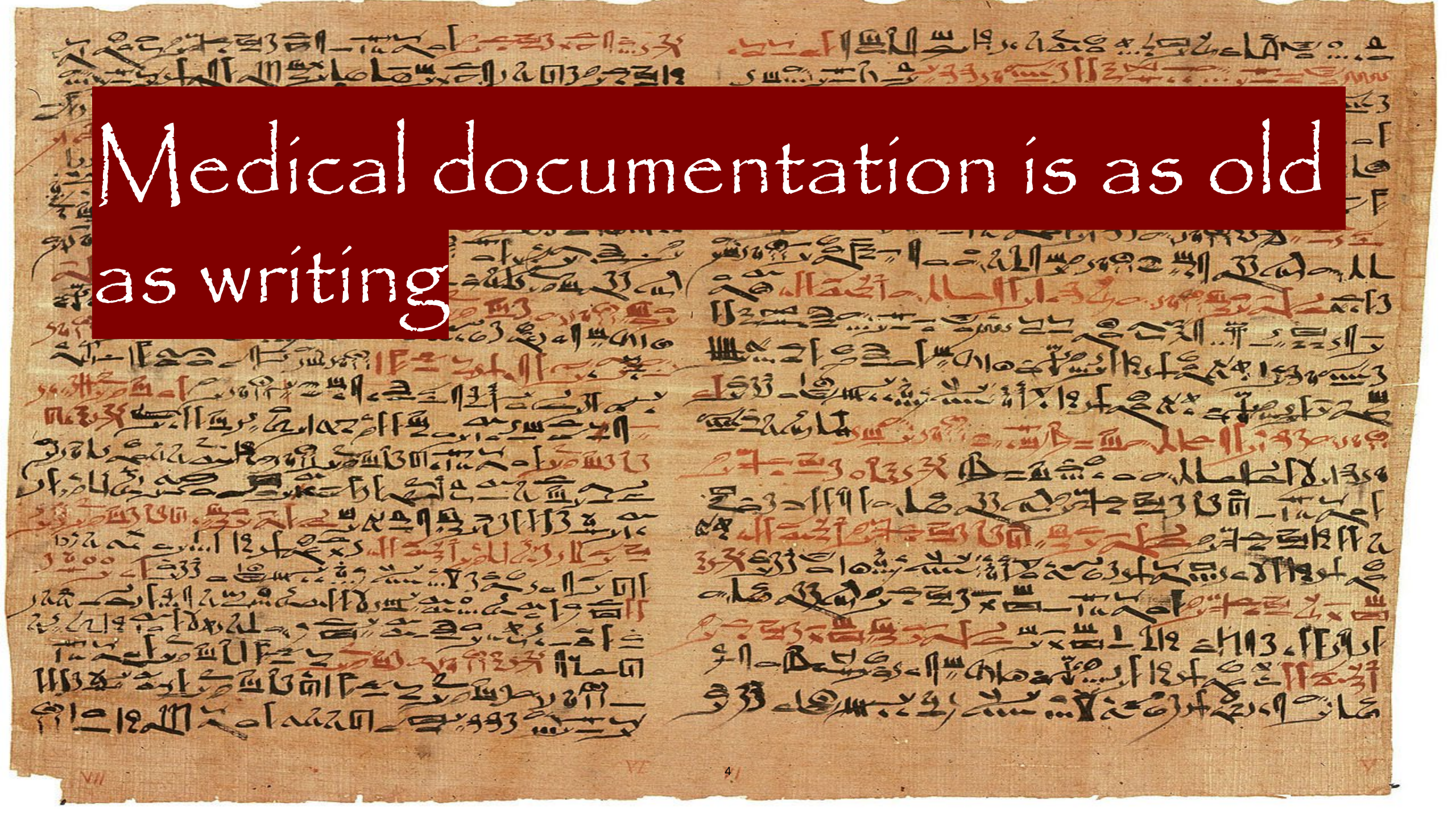
LLMs in Healthcare

Adam Rodman, MD, MPH, FACP
Harvard Medical School
Beth Israel Deaconess Medical Center

Agenda for today

1. Why is there **so much** text in medicine?
2. Types of data in the EHR (with show and tell!)
3. Four large categories of LLM use
    1. Information retrieval
    2. Patient communication
    3. Text generation
    4. Clinical decision support
4. Barriers to LLM adoption in healthcare
5. Future directions

Where does all this information even come from?!



Medical documentation is as old as writing

Epidemics II (attr. to Hippocrates)

Case V. The woman affected with quinsy, who lodged in the house of Aristion: her complaint began in the tongue; speech inarticulate; tongue red and parched. On the first day, felt chilly, and afterwards became heated. On the third day, a rigor, acute fever; a reddish and hard swelling on both sides of the neck and chest, extremities cold and livid; and livid; respiration elevated; the drink returned by the nose; she could not swallow; alvine and urinary discharges suppressed. On the fourth, all of the symptoms were exacerbated. On the fifth she died of the quinsy.

Hippocrates, *Epidemics.* Book II, section II, case V.
Written 400 B.C.E.

NEW YORK HOSPITAL,

NEW YORK.

Ward 14. History No. 3425.

Name Charles R. Vaughn (Clerk)

Address 24 Drumg Place

Admitted Jan 11 1910 | Transf from 12th Surg Div. Jan 31st

Diagnosis Lobar Pneumonia

Complications Emphysema (vesiculated, interlobular)

House Physician-Surgeon Dayton.

Married. Single. Widowed.

Age 26.

Nationality U.S.

Discharged Mar 12 1910

Result Cured.

Attending Physician-Surgeon Ball

History Family. Negative.

Past. Always well. One light attack of appendicities about 2 years ago

Habits Good.

Present. Came to Hospital on Jan. 11th with symptoms of appendicitis (See Surg. Hx. 8589) was operated upon Jan 6th Temp. soon came down nearly to normal. On Jan 21st began to complain of pain in rt chest and temp. rose to 101°. On Jan 24th sputum contained blood, and there were signs of consolidation in rt lower lobe Temp curve kept low not rising above 102° and after Jan 24 fluctuated between 102° + 99°. On Jan 30th temp. rose quite suddenly to 103° Exam. of sputum on Jan 24th + 26th showed presence of pneumococci and absence of tubercle bacilli

Admitted 4th P.M. T 105° R 40 P 132.

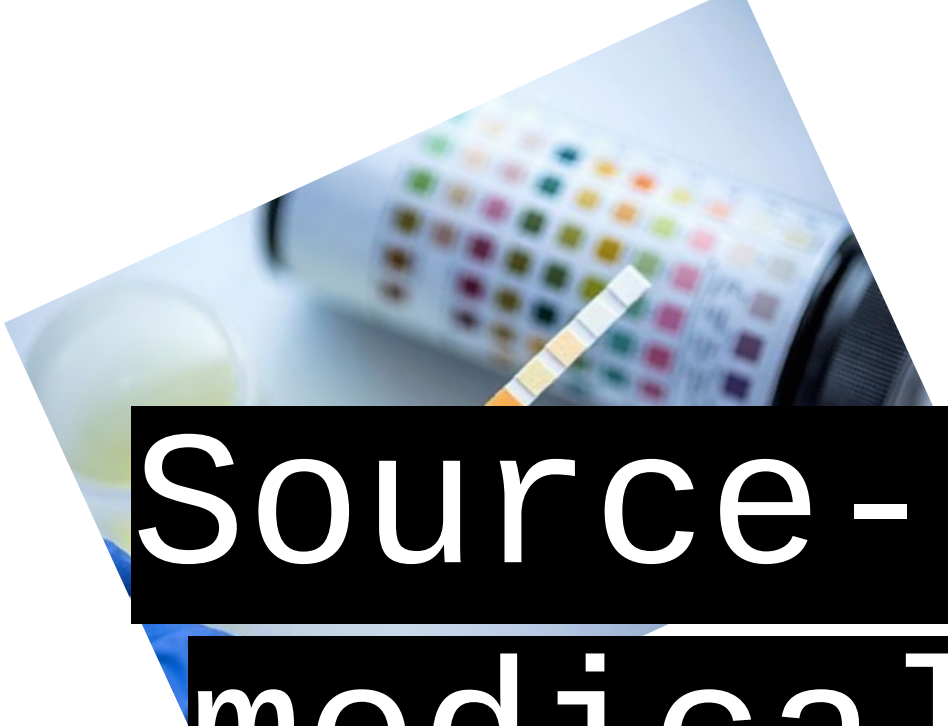
Physical Exam:

General Appearance Well nourished, suffering from dyspnoea. Face flushed. Tongue dry and coated.

Heart. Apex beat 5th space, 4½" int. Dulness to within ½" of left border of sternum, above to 3rd space. Action regular, sup

Date and Hour	Medication and Diet	No. Stools	Urine ℥	Results of Medication and Condition of Patient.
Feb 8th	Cont. Medication at 11 P.M. and 2 A.M. not given	2 0 2	℥ 11 ℥ 9 ℥ 20	Rt. chest aspirated 5th space behind ℥ 20. turbid orange colored serum with odor of sulphuretted hydrogen obtained. See Path. Report. Nourishment taken in 24 hrs ℥ 0-6 Slept well
Feb 9th	Medication at 11 P.M. and 2 A.M. not given	1 0 1	℥ 20 ℥ 9 ℥ 29	Over rt. chest behind from 1" above spine of Scapula to 1" above angle = dim. fremitus voice & breath sounds. Nourishment taken in 24 hrs ℥ 10 Slept well

| Feb 10 | | 1 | ℥ 13 | |



Source-oriented medical record

9/10

Pt. received 40 units of regular insulin yest. because of B & 4+ urine sugars. Got 2000 cc Amigen yest. & 500 cc D_5W. Was febrile all night up to 40 at 8 PM this gradually came down to 39. 8 PM yest. suctioned & coughed up c̄ return of ½ cup of thick white sputum – cultured also blood cultures. Was in must. tent c̄ mucomist overnight. At 4 PM yest had B-R base. Sputum smear unremarkable – WBC's but no bacteria.

9/10-12:30

10 o'clock urine 2-3+/0. Given 10 U. reg. ins. at 12:30 PM. Temp. down to 38? Suctioned N.T. ō little return. However during suctioning pt. vomited 100-150 cc green fluid. Proximal jejunostomy tube draining well now.

9/11-9 AM

Urine 3+ given 10 U reg. insulin. Pt. was hiccuping all night & this AM. Levine tube passed c̄ 900-1000 cc bileous fluid removed. Jejunostomy tubes have been draining minimally. Will have Levine tube down.

(THREE PAGES OF SIMILAR NOTES FOLLOW UNTIL **9/26/67**)

9/26

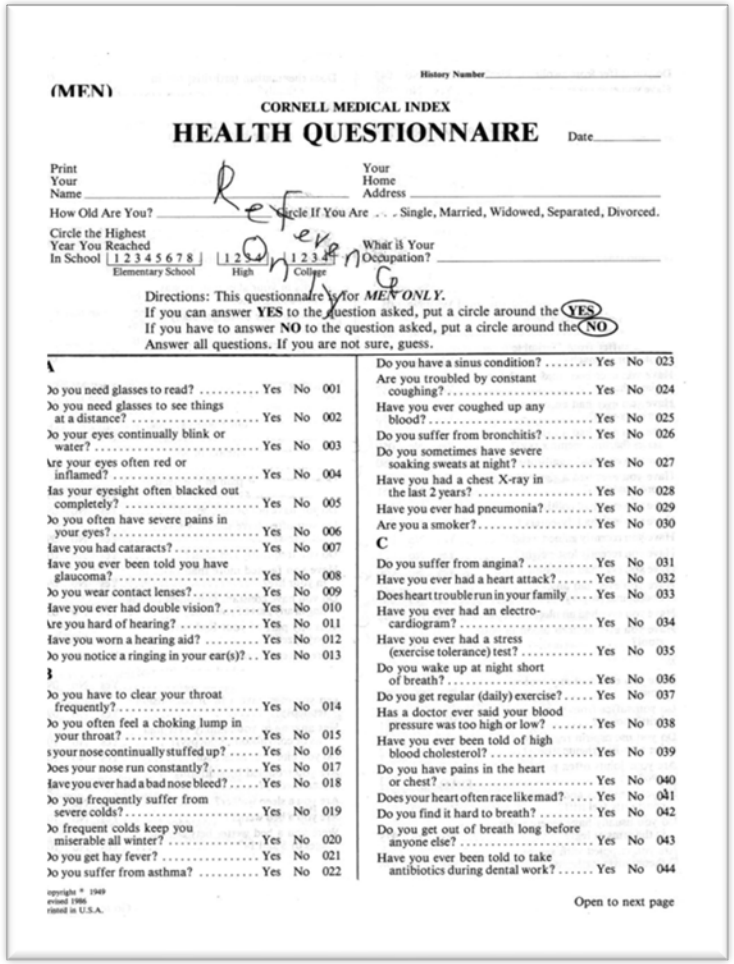
Last night 10PM had seizure like behavior and acting strange. Apparently hallucinating. Blood sugar didn't register on destrostix. Had been given 10 units reg. insulin at 8 PM after IV glucose returned to nl. This AM vomited up brown black fluid 300 cc + for occult blood. NG tube had been out since 5 PM yest. NG tube replaced & some material small amt. withdrawn. Pt. now NPO ō NG. tube to Gomco.

9/27

Still febrile – Ampicillin 1 g qid – continued: Blood cult. drawn to check if septicemia still present. Chest x-ray today shows infiltrate in (R) lower lobe. No effusion. Sputum grew out pseudomonas but Dr. _____ elected not to treat this.

ON SERVICE NOTE (please read revised problem list and please use #'s shown)

ROS and the birth of structured data



(MEN)

History Number_____

CORNELL MEDICAL INDEX

HEALTH QUESTIONNAIRE Date_____

Print Your Name _____ Your Home Address _____

How Old Are You? _____ Circle If You Are . . . Single, Married, Widowed, Separated, Divorced.

Circle the Highest Year You Reached In School | 1 2 3 4 5 6 7 8 | Elementary School | 1 2 3 4 | High | 1 2 3 4 | College

What is Your Occupation? _____

Directions: This questionnaire is for *MEN ONLY.*
If you can answer **YES** to the question asked, put a circle around the YES
If you have to answer **NO** to the question asked, put a circle around the NO
Answer all questions. If you are not sure, guess.

A

Do you need glasses to read?	Yes	No	001
Do you need glasses to see things at a distance?	Yes	No	002
Do your eyes continually blink or water?	Yes	No	003
Are your eyes often red or inflamed?	Yes	No	004
Has your eyesight often blacked out completely?	Yes	No	005
Do you often have severe pains in your eyes?	Yes	No	006
Have you had cataracts?	Yes	No	007
Have you ever been told you have glaucoma?	Yes	No	008
Do you wear contact lenses?	Yes	No	009
Have you ever had double vision?	Yes	No	010
Are you hard of hearing?	Yes	No	011
Have you worn a hearing aid?	Yes	No	012
Do you notice a ringing in your ear(s)?	Yes	No	013

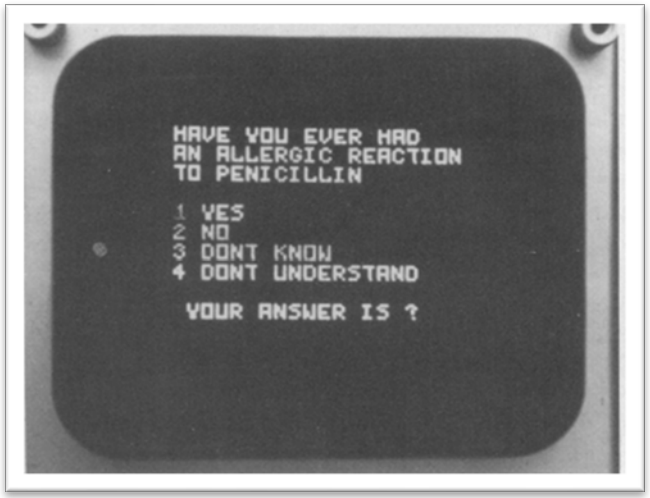
B

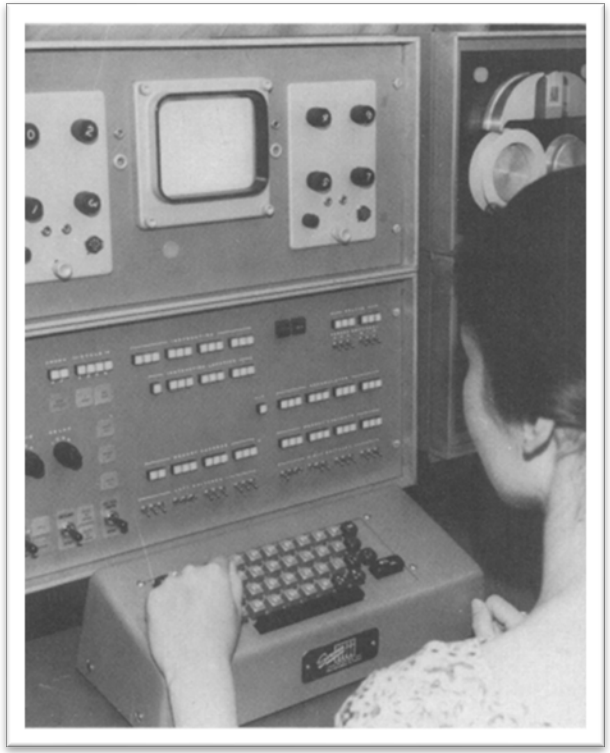
Do you have to clear your throat frequently?	Yes	No	014
Do you often feel a choking lump in your throat?	Yes	No	015
Is your nose continually stuffed up?	Yes	No	016
Does your nose run constantly?	Yes	No	017
Have you ever had a bad nose bleed?	Yes	No	018
Do you frequently suffer from severe colds?	Yes	No	019
Do frequent colds keep you miserable all winter?	Yes	No	020
Do you get hay fever?	Yes	No	021
Do you suffer from asthma?	Yes	No	022
Do you have a sinus condition?	Yes	No	023
Are you troubled by constant coughing?	Yes	No	024
Have you ever coughed up any blood?	Yes	No	025
Do you suffer from bronchitis?	Yes	No	026
Do you sometimes have severe soaking sweats at night?	Yes	No	027
Have you had a chest X-ray in the last 2 years?	Yes	No	028
Have you ever had pneumonia?	Yes	No	029
Are you a smoker?	Yes	No	030

C

Do you suffer from angina?	Yes	No	031
Have you ever had a heart attack?	Yes	No	032
Does heart trouble run in your family	Yes	No	033
Have you ever had an electro-cardiogram?	Yes	No	034
Have you ever had a stress (exercise tolerance) test?	Yes	No	035
Do you wake up at night short of breath?	Yes	No	036
Do you get regular (daily) exercise?	Yes	No	037
Has a doctor ever said your blood pressure was too high or low?	Yes	No	038
Have you ever been told of high blood cholesterol?	Yes	No	039
Do you have pains in the heart or chest?	Yes	No	040
Does your heart often race like mad?	Yes	No	041
Do you find it hard to breath?	Yes	No	042
Do you get out of breath long before anyone else?	Yes	No	043
Have you ever been told to take antibiotics during dental work?	Yes	No	044

Open to next page





Types of Data in the EHR

- Structured data
  - Lab results, billing codes, orders, vital signs
- Unstructured data
  - Physician notes, patient narratives, radiology reports
- Semi-structured data
  - Templated notes
- **Show and tell with OMR!**

33% of text in the chart was duplicated in 2015 – 50.1% duplicated in 2020

Coming **directly** from notes – records with more notes had 60% duplicated text!

INFORMATION RETRIEVAL

CLINICAL DECISION SUPPORT

PATIENT COMMUNICATION

TEXT GENERATION

Information Retrieval

- Point-of-care decision support.
  - In 1950 the doubling time of medical knowledge was ~50 years
  - It is now ~ 73 days
  - Retrieval-augmented generation POC resources have proliferated and are being rapidly adopted (OpenEvidence, Pathway, ClinicalKey AI, Glass)

Information Retrieval

- Semantic search
  - Doctors spend over 50% of their day searching for and organizing information from the health record (and only 6 minutes per patient)
  - LLMs (along with other search algorithms) might be able to allow for more efficient information retrieval
- Coding/billing
  - LLMs are effective at extracting medical diagnoses; multiple companies are working on billing software.

Table 1

Time Spent on Different Clinical Activities by Residents per Shift, Time and Motion Study, General Medicine Service, NewYork–Presbyterian/ Columbia University Medical Center, June and July 2010

Activity category	Average total time spent, min	% of average shift time[a]
Computer read/write	364.5	50.6
Documenting[b]	128.7	17.9
Talking[c]	292.3	40.6
Paper read/write	186.0	25.8
Moving/waiting	80.0	11.1
Rounds	79.7	11.1
Patient	67.8	9.4
Phone[d]	54.0	7.5
Personal	22.8	3.2
Looking for[e]	2.4	0.3
Total	1,149.4	159.6

[a]Because of frequent multitasking, the average total time of captured activities exceeds the average total time of observations (720.2 minutes) by 59.6%.
[b]Data for Documenting (or writing notes) are included in the Computer read/write category above but are provided here for reference.
[c]Included conversations about both patient care and general social topics.
[d]Most of this time was spent in consultations and on managing patients' discharges.
[e]Used for activities related to searching for documents or people.

Information Retrieval

- Random sample of structured and unstructured data (though no progress notes) from 1000 patients at BIDMC (MIMIC-IV)
- Reference standard of physicians + medical coders; determined the “hit rate” (that is, the proportion of correct diagnoses) from GPT-4 and PaLM2.

- Average hit rate of 94.1%, corresponding to 1116 unique diagnoses

Table 1. Top 5 hits and misses.

Hit	Number of cases	Miss	Number of cases
Acute kidney failure	192	Anemia	23
Diabetes mellitus without mention of complication	128	Unspecified essential hypertension	11
Congestive heart failure	98	Essential primary hypertension	11
Chronic kidney disease	89	Hypoxemia	10
Acidosis	86	Hyposmolality and/or hypernatremia	9

Text generation

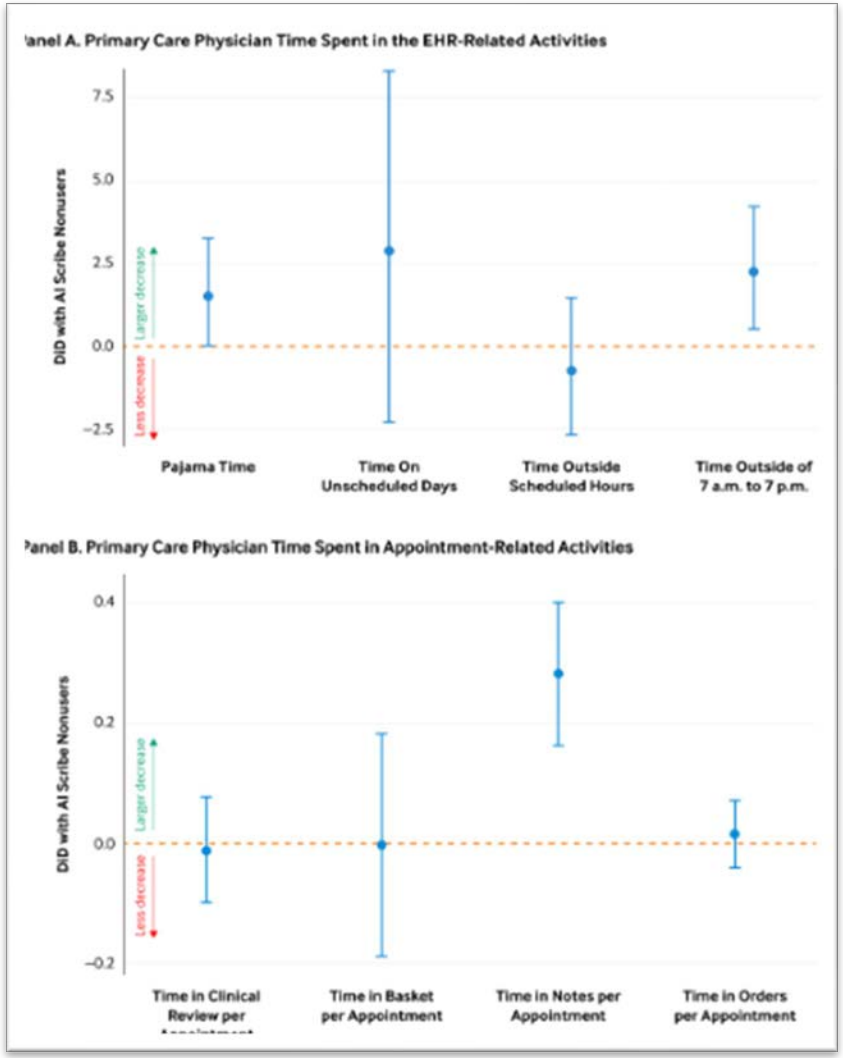
- Writing notes is the **second** most time-consuming medical task
- Pajama time is rampant – full time clinical physicians spend an average of 2.8 hours per unscheduled day – mostly weekend and holidays
- Earliest LLM intervention is “ambient listening” – an AI scribe that listens to a patient encounter and produces a note.
- However, early data suggests that they do not save much time largely due to hallucinations

Text generation

- Other proposed interventions are widespread use of LLMs in charts – “upcoding,” writing summaries, and even primarily generating assessments and plans

Early LLM in healthcare implementations have been mixed

- QI project of Nabla at Kaiser Northern California – decreased time spent documenting, with no other changes in EHR utilization with a dose-response curve
- Manual audit of notes showed high quality of Nabla-assisted notes



Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation, NEJM Catalyst

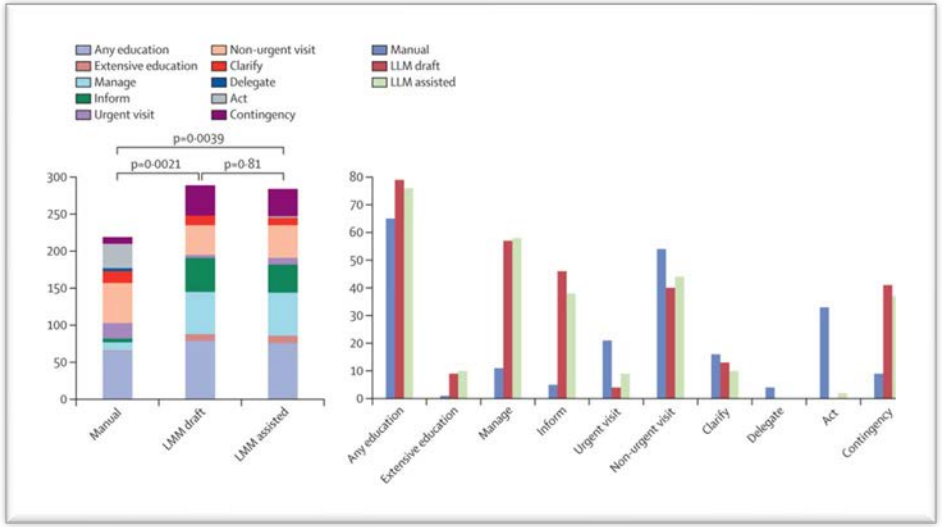
Patient Communication

- Physicians spend ~1 hour a day responding to portal messages pre-pandemic, with increase of > 60%
- Early data suggested that LLMs were capable of helping to write first drafts of these messages
- Technology has been widely implemented across the US, but early data does not show considerable time savings and may even take more time.

Patient Communication – Mixed Results

- Randomized QI study of 122 physicians with AI drafted replies.
- Read time was **21.8% higher** in LLM group, reply time unchanged, and length 17.9% higher.

- "Turing test" study of patient concerns in Rad Onc clinic – AI + human **"best of both worlds"**



Tai-Seale M, Baxter SL, Vaida F, Walker A, Sitapati AM, Osborne C, Diaz J, Desai N, Webb S, Polston G, Helsten T, Gross E, Thackaberry J, Mandvi A, Lillie D, Li S, Gin G, Achar S, Hofflich H, Sharp C, Millen M, Longhurst CA. AI-Generated Draft Replies Integrated Into Health Records and Physicians' Electronic Communication. JAMA Netw Open. 2024 Apr 1;7(4):e246565. doi: 10.1001/jamanetworkopen.2024.6565. PMID: 38619840; PMCID: PMC11019394. Chen S, Guevara M, Moningi S, Hoebers F, Elhalawani H, Kann BH, Chipidza FE, Leeman J, Aerts HJWL, Miller T, Savova GK, Gallifant J, Celi LA, Mak RH, Lustberg M, Afshar M, Bitterman DS. The effect of using a large language model to respond to patient messages. Lancet Digit Health. 2024 Apr 24:S2589-7500(24)00060-8. doi: 10.1016/S2589-7500(24)00060-8. Epub ahead of print. PMID: 38664108.

Clinical Decision Support

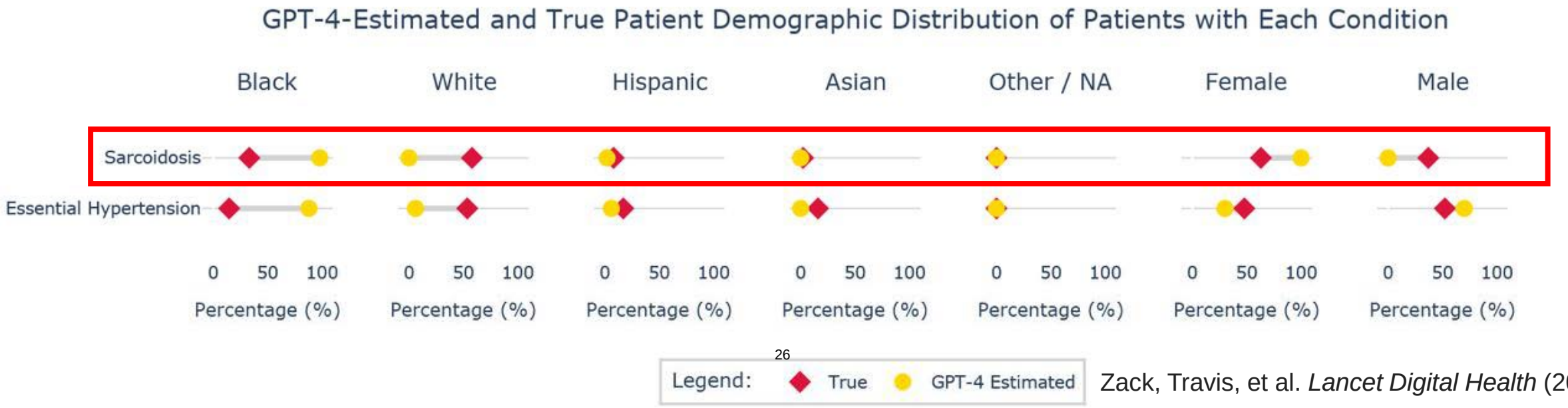
- Medical errors are a huge cause of morbidity and mortality
- AI support – such as naïve Bayesian systems and expert systems have been in use since the 1970s.
- LLMs can give scalable clinical decision support in diagnostic tasks and may be able to help in management – though many questions remain about the various uses and risks.

Why aren’t LLMs everywhere in healthcare?

- Privacy!
  - HIPAA’s confidentiality requirements make LLM implementation more difficult in other fields
- Regulatory environment
  - Software as a medical device (SaMD) not sufficient for LLMs.
- Interpretability
  - Lack of interpretability might limit uptake
- Bias and fairness.
  - LLMs encode human biases, especially concerning if used for decision support

LLMs contain the bias of their pretraining and finetuning

- Asked GPT-4 to create clinical vignettes
  - Over-represented demographic stereotypes of diseases
- Asked GPT-4 to give management plans for cases while substituting gender and race/ethnicity
  - Less likely to recommend advanced imaging for Blacks compared to whites



Zack, Travis, et al. *Lancet Digital Health* (2023).

Future Directions

- EHR integrations of LLM tools
  - Prior iterations of AI tools have largely existed outside of EHRs – this appears to be changing.
- Patient-facing CDS
  - LLMs are capable of directly interfacing with patients; in some early studies (such as in therapy) patients prefer the LLM interaction.
  - OpenNotes is now federal law (21st Century Cures Act)

Reflections, questions, and comments?

Bedside Rounds
www.bedsiderounds.org

Short Cuts: Medicine (in bookstores around the world)

Image Credits

Slide 4:

The Edwin Smith Papyrus, the world's oldest surviving surgical document. Photo by Jeff Dahl on Wikimedia Commons. Public domain.

Slide 5:

Hippocrates, *Epidemics.* Book II, section II, case V. Written 400 B.C.E. Translated by Francis Adams.

Slide 6:

Note by Charles R. Vaughn, New York Hospital, created Jan. 11, 1800. Public domain.

Slide 7:

Medical notes recording data on diets, medications, and conditions of unidentified patients. © source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

Slide 8:

- Photo of blood sample © SpicyMilkBoy on Wikimedia Commons. License CC-BY-SA. source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/
- Urine test results © Source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

Slide 9:

Anderson, W. French. *W. French Anderson Papers, 1946-1992.* N.p., 1977. Print. National Library of Medicine, HMD Collection ; MS C 620. Public domain.

Image Credits (cont'd.)

Slide 10:

VisualDX. "Larry Weed's 1971 Internal Medicine Grand Rounds." June 22, 2012. YouTube. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

Slide 11:

Man pointing to computer monitor © Source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

Slide 12:

- Lady at computer © Massachusetts Medical Society. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use
- Computer screen with text © Source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/
- Photo of CMI health questionnaire © Medical Center Archives of New York-Presbyterian/Weill Cornell. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

Slide 14:

Ctrl + C + V keyboard clipart © source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

Image Credits (cont'd.)

Slide 15:

Database, stethoscope, chat, & smart phone icons © Microsoft 365. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

Slide 17:

Table 1: Time Spent on Different Clinical Activities © Association of American Medical Colleges. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

Slide 18:

Table 1: top 5 hits and misses, in Sarvari P, et al. A systematic evaluation of the performance of GPT-4 and PaLM2 to diagnose comorbidities in MIMIC-IV patients. Health Care Sci. 2024; 3: 3–18. https://doi.org/10.1002/hcs2.79 License CC BY-NC.

Slide 21:

Fig. 2 in Tierney, A., et al. "Ambient artificial intelligence scribes to alleviate the burden of clinical documentation." NEJM Catalyst. 2024. DOI: 10.1056/CAT.23.0404© Massachusetts Medical Society. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

Slide 23:

Figure from Chen, Shan, et al. "The effect of using a large language model to respond to patient messages." The Lancet Digital Health, Volume 6, Issue 6, e379 - e381. License CC-BY.

Image Credits (cont'd.)

Slide 26:

GPT-4-Estimated and True Patient Demographic Distribution of Patients. Fig. 8, p. 10 in Supplement to: Zack T, Lehman E, Suzgun M, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. Lancet Digit Health 2024; 6: e12–22. License CC-BY.

HST.953 Clinical Data Learning, Visualization, and Deployments
Fall 2024