RICE

Bridging the Information Gap: Utility and Applications of NLP in Healthcare Applications

Andrew Wen

April 01, 2026

… Presented from the Perspective of a *Clinical Translational* Data Scientist

- How do we bring the latest advances in basic science to real-world applications?
- Focus on AIML to accelerate clinical research and support clinical practice using data derived from ***clinical care***
- Discussion today will be through that lens. Notably, this excludes Bioinformatics/Computational Genomics, Biomedical Engineering, and Consumer Health
- *Application* Rather than *Methodological* Focus

About Me

- Clinical Data Scientist with a Focus on Natural Language Processing-Based Analytics and Applications on Big Data (Data Mining, Information Retrieval) for Clinical Decision Support
- 2015-2017: RA, OHSU School of Medicine, Department of Medical Informatics and Clinical Epidemiology
  - Enhancing Vector Space Model-based Information Retrieval with NLP Artifacts
- 2017-2023: Senior Data Science Analyst, Mayo Clinic, Department of AI & Informatics Research
  - Multimodal Information Retrieval, Text2Query for Clinical Trial Matching, Human-in-the-loop Information Extraction, Clinical Text Deidentification, NLP for Clinical Information Exchange
- 2023-2026: Data Scientist II, UTHealth Houston, Center for Translational AI Excellence and Applications in Medicine
  - Clinical Association Mining and Knowledge Graph for Algorithm Distillation, Multi-agent Human-in-the-Loop Information Retrieval and Data Mining, Data Mining from Unstructured Big Data

- Opinions my own, and does not necessarily represent the official views of the University of Texas System or the State of Texas

An Estimated 70-80% of Clinically Relevant Information is Found Only in Unstructured Data

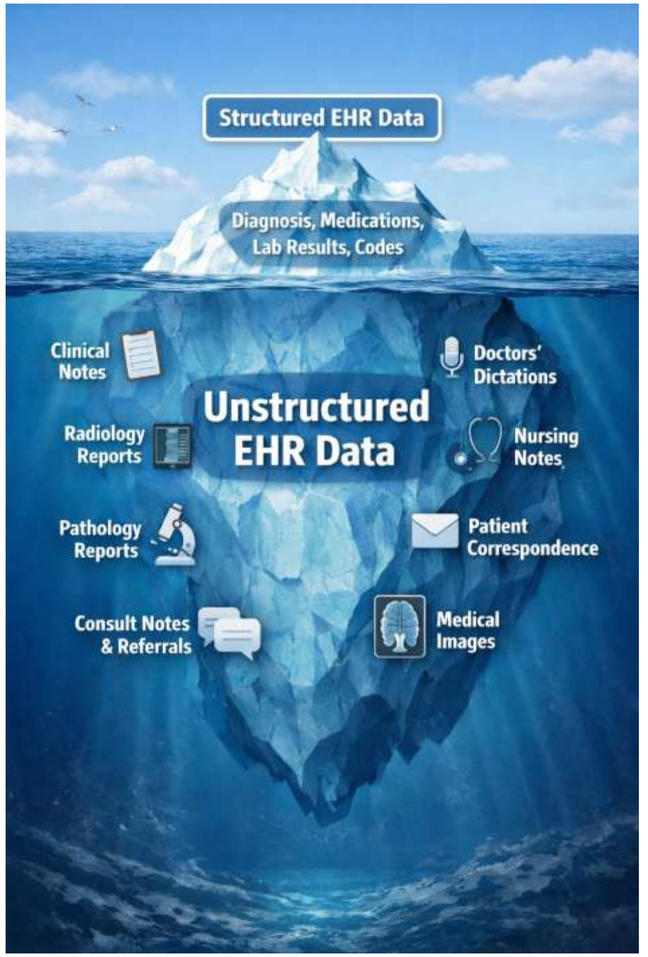
… and due to clinical interpretation, text by proxy

The Intent Behind Which Data is Collected and Generated Matters

- Structured -> Insurance claims and associated justification
- An accurate (but incomplete!) view of a patient's clinical history
- Debatably focused around revenue maximization
- *Secondary* use of clinical data (HITECH act of 2009)

Substantial clinically-relevant information not captured:

- (In general) Signs/Symptoms
- Outside Materials/Cross-Institutional Transfer of Information
- *Clinical Interpretation of Raw Information*
- *Clinical Reasoning for Decision-Making*



RICE

Structured Data Expresses the *What*, Text Data Expresses The Reasoning as to *Why*

Assessment/Plan

1. Chest pain. At this point, because of Dr. XXX's evaluation last year and the symptoms exactly the same, I think this is noncardiac. My intonation is that this is reflux. I am going have her double her Aciphex or increase it to b.i.d., and I am going to have her see Dr. XXX for an EGD. She is to let me know if her symptoms are getting worse or if she is having any severe episodes.
2. Suspicious Lesions on Left Shoulder: We will do a punch biopsy and set her up for an appointment for that.
3. History of leukopenia. We will check a CBC.

Patient ID	Condition	Date
12345	GERD	2026-06-01
12345	Skin Lesions	2026-04-01
12345	Leukopenia	2021-01-01

Patient ID	Procedures	Date
12345	EGD	2026-06-01
12345	Punch Biopsy	2026-05-01

Patient ID	Medications	Date
12345	Aciphex	2026-04-01

Patient ID	Lab Tests	Date
12345	CBC	2026-04-01

Clinical Language is Fundamentally Challenging to Parse

Lexical Semantics:

Patient is a 53 y/o male presenting for follow up on his bone CA and is being prepped for discharge once post-surgical discharge subsides.

- Homonymy: Is this “cancer”, “calcium”, or “California”?
- Polysemy: Release from hospital vs. fluid “released” from surgical wound
- Also: Synonymy (Particularly Abbreviations), but also consider “Cephalgia” vs. “Headache”

Context Cues:

Patient is a 23 y/o male presenting w/ acute chest pain. DDx acute coronary syndrome, pulmonary embolism. R/o for aortic dissection and GERD. PHx MI, T2D. FHx Gastric CA.

- Personal History vs. Family History: Patient vs. Family Member has Condition. Neither of these conditions are currently active.
- Hypotheticals: Uncertain which one of the two the patient has and are positing that it is one of the two
- Negation: “Ruled out for” we previously posited these hypotheticals, and subsequently ruled them out

- *Recall that context window lengths are limited and the vanishing gradient problem!*

Applications of Natural Language Processing in Healthcare

The Role of NLP In Healthcare

- The Information Gap
- Clinical Reasoning vs. Facts
- Homonymy, Polysymy, Synonymy, and Context

The Foundations: Applications of Natural Language Processing in Support of Clinical Research

- In-Silico Clinical Phenotyping for Cohort Accrual and Recruitment
- PubMed and Evidence-Based Medicine
- Digital Twins and “Patients Like Me”

The Applications: Clinical Natural Language Processing supporting Healthcare AIML Translation for Clinical Practice

- Clinical Decision Support on Real World Data
- Formalization of Clinical Knowledge via Health Knowledge Graphs and Ontologies
- Knowledge Distillation for Low-Resource AI

The Frontiers: Clinical AIML in the Post-Generative Language Model Era

- Ambient Scribes and Encounter Summarization
- Patient Portals, Provider-Patient Messaging, and Telemedicine

The Foundations: Applications of NLP in Support of Clinical Research

RICE

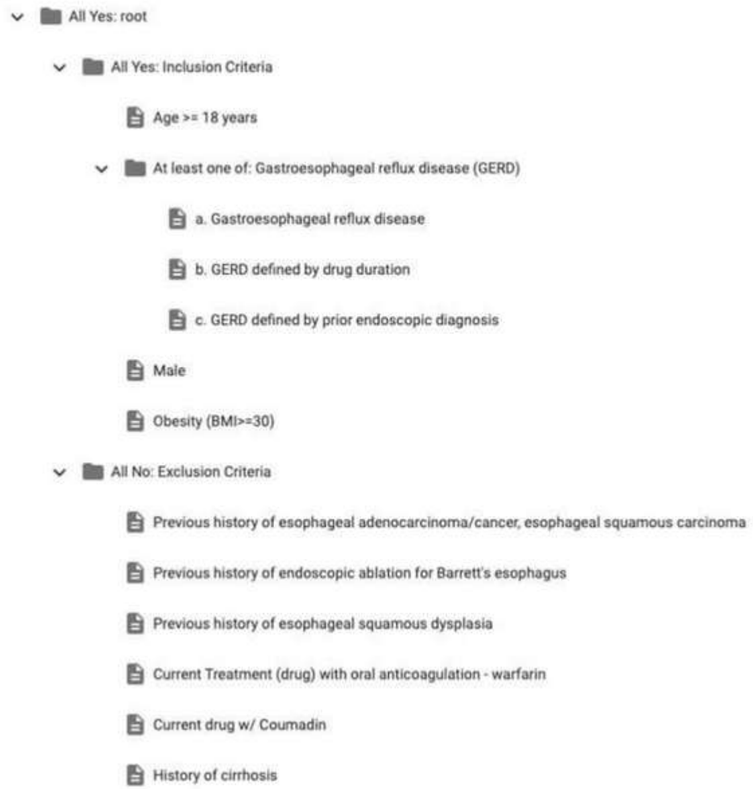
Use Case 1: Defining patient cohorts is central for clinical research

- Typically established via a set of clinical criteria to define case/control groups

- Used for a variety of clinical research projects
  - Longitudinal Cohort/Epidemiological Studies
  - Clinical Trials
  - “Labeled” Data for AIML Studies

- *Often the bottlenecking step of any clinical research study*

- Example: I want to recruit patients with GERD

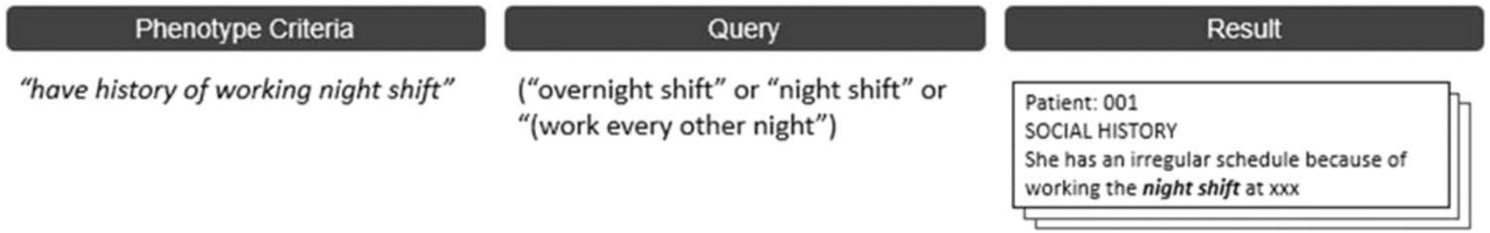


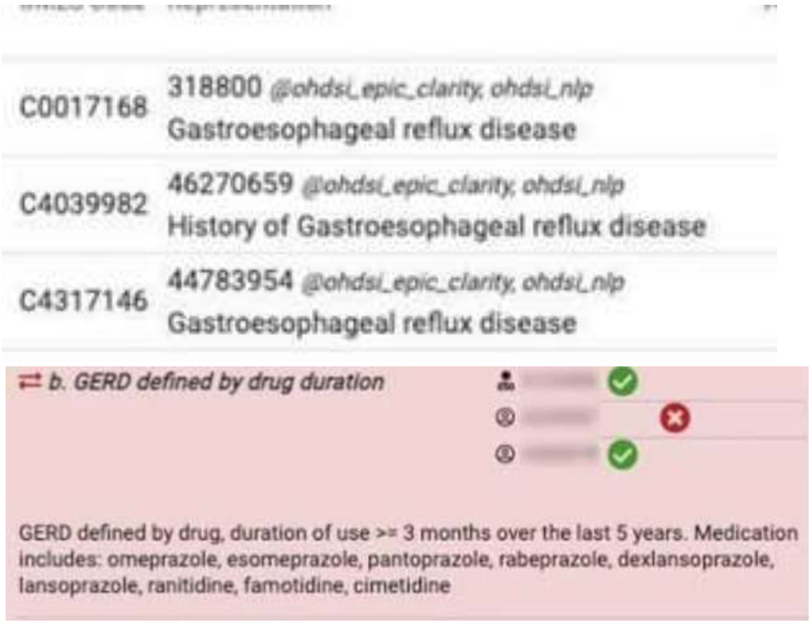
Source: ***Wen, A.***, *He, H., Fu, S. et al. The IMPACT framework and implementation for accessible in silico clinical phenotyping in the digital era. npj Digit. Med. 6, 132 (2023). https://www.nature.com/articles/s41746-023-00878-9*

Use Case 1: Clinical phenotyping is difficult due to incomplete, noisy, and heterogeneous data

Requiring Clinical Definitions and Interpretation and Cross-Modal Integration

- Information Needs can Span a Variety of Modalities and Cross-Modality Information Often Contradicts w/ Varying Representations
- Expansion of Clinical Definitions Often Requires [Multiple Rounds of] Clinical Engagement





Source: **Wen, A.**, *He, H., Fu, S. et al. The IMPACT framework and implementation for accessible in silico clinical phenotyping in the digital era. npj Digit. Med. 6, 132 (2023). https://www.nature.com/articles/s41746-023-00878-9*

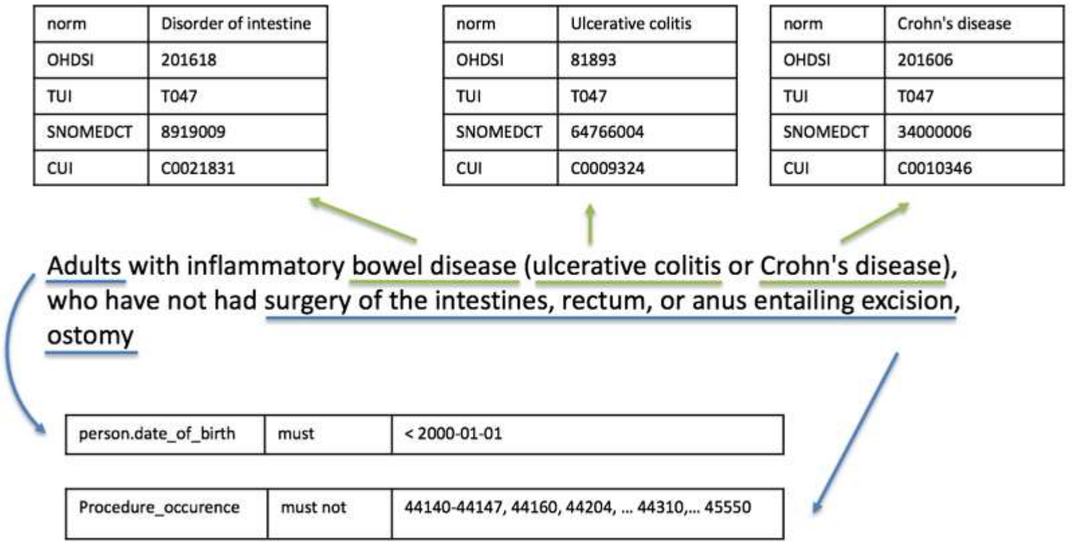
RICE

Use Case 1: Clinical phenotyping inherently requires human-in-the-loop clinical domain expertise

- It is precisely this multi-turn engagement between data engineers and clinical domain experts that leads to clinical phenotyping being a many-months-long process

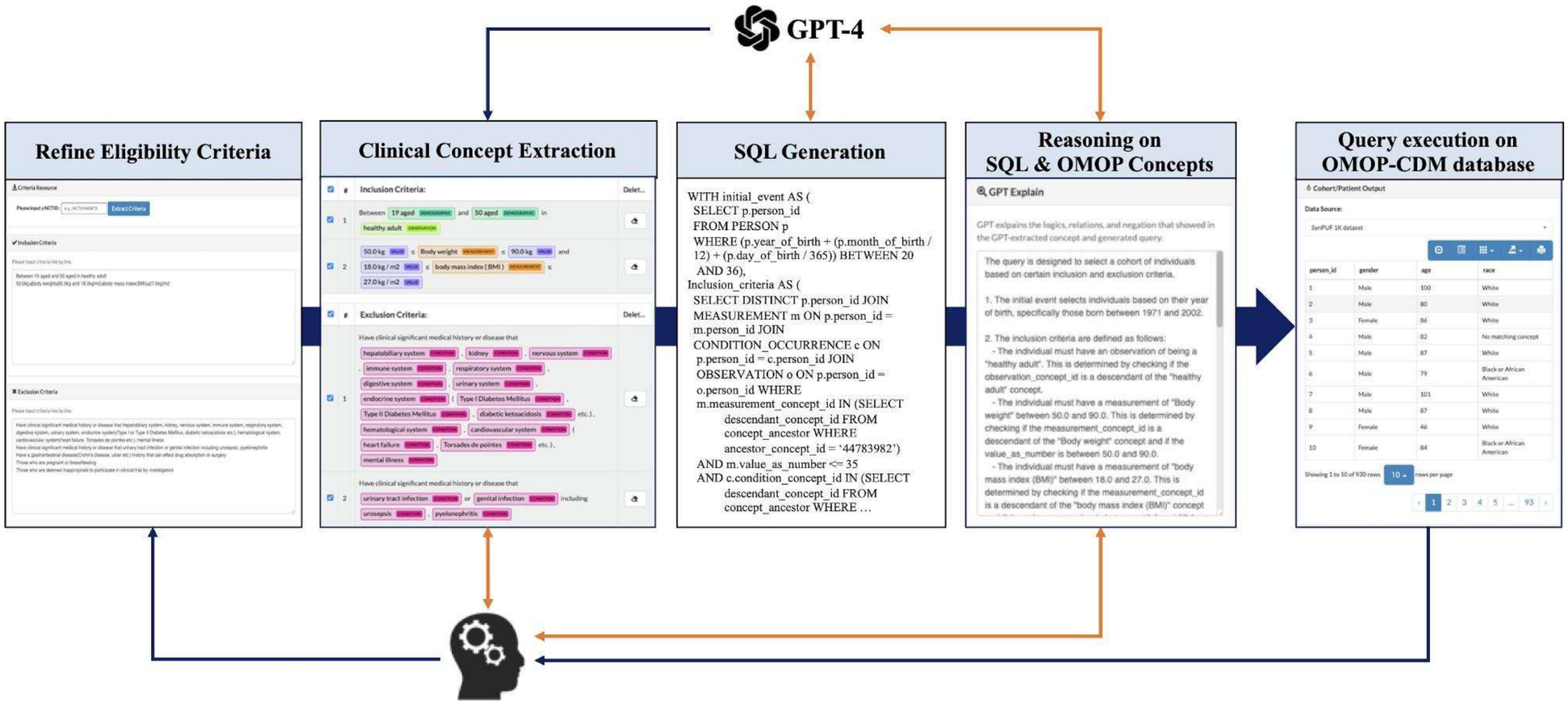
- (Most) clinicians do not have the domain expertise to define where certain bits of information are located within a clinical data warehouse, nor do they know how to query it.

- (Most) data engineers can tell you where information is located within a clinical data warehouse, but cannot tell you if your retrieved results match a given clinical criteria



Source: *Liu S, Wang Y, **Wen A**, Wang L, Hong N, Shen F, Bedrick S, Hersh W, Liu H*
Implementation of a Cohort Retrieval System for Clinical Data Repositories Using the Observational Medical Outcomes Partnership Common Data Model: Proof-of-Concept System Validation
JMIR Med Inform 2020;8(10):e17376

Use Case 1: NLP can be used to support human-in-the-loop text2query



Park J, Fang Y, Ta C, Zhang G, Idnay B, Chen F, Feng D, Shyu R, Gordon ER, Spotnitz M, Weng C. Criteria2Query 3.0: Leveraging generative large language models for clinical trial eligibility query generation. Journal of biomedical informatics. 2024 Jun 1;154:104649.

RICE

Use Case 2: Clinicians and researchers must navigate an overwhelming volume of biomedical literature

Evidence-Based Medicine: a paradigm of thought in clinical care where treatment decisions are made via clinicians combining what is observed about a patient alongside current best evidence in research (e.g., found through PubMed) and their own clinical experience.

- EBM has dominated clinical care practice in the past thirty years
- If known, apply knowledge, if unknown, look it up!

Knowing what is unknown and looking it up is the tricky part!

- There are multi-billion dollar industries dedicated purely to facilitating this problem (e.g., UpToDate, OpenEvidence)

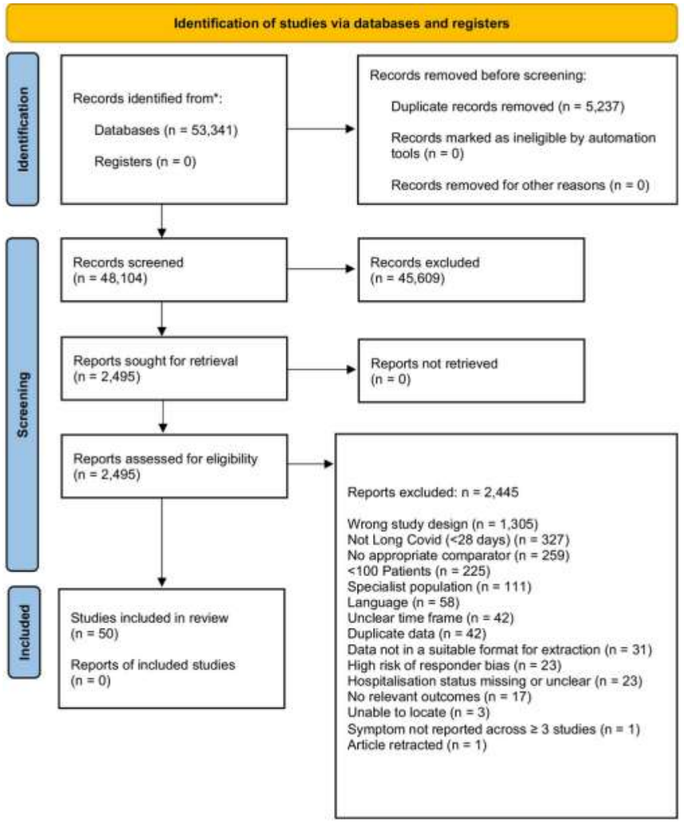
	FY2023
MEDLINE Citations Indexed (Annual)	1,279,327
MEDLINE Citations Cumulative Total	30,966,708
MEDLINE Journal Titles	5,294
PubMed Citations (Annual)	1,567,478
PubMed Citations Cumulative Total	36,555,430

Source: https://www.nlm.nih.gov/bsd/medline_pubmed_production_stats.html

Use Case 2: Biomedical search is a semantic retrieval problem under vocabulary mismatch

- Recall issues of polysemy, homonymy, synonymy, and context
- Systematic Reviews the “Gold Standard of Evidence”
  - Very costly and time-consuming, but needs to be a convincing argument that it is a ***comprehensive*** overview of current state of the field

String	Search Terms
COVID terms	covid* or ncov* or novel coronavirus or novel betacoronavirus or sars-ncov-2 or sars-cov-2
Any of these within three words of any of the COVID terms	Long, chronic, long term, long-term, longterm, sequela*, post acute, post-acute, postacute, long haul*, long-haul*, longhaul*, survivor*, on going, on-going, ongoing
Any of these within five words of any of the COVID terms	persist*, recover*, discharg*, follow*, prolong*
Any of these combinations	post-covid syndrome, postcovid syndrome, post covid syndrome symptomatic within three words of any of the COVID terms, but also with chronic, and any of long term, long-term, longterm Any of chronic*, long term, long-term, longterm within three words of any of post covid, post-covid, postcovid



Source: *O’Mahoney, L.L., Routen, A., Gillies, C. et al. The risk of Long Covid symptoms: a systematic review and meta-analysis of controlled studies. Nat Commun **16**, 4249 (2025). https://doi.org/10.1038/s41467-025-59012-w*

RICE

Use Case 2: The rise of NLP for secondary relevance screening

- Pure TF-IDF-based methods fast, but either trade recall or precision.
  - Just think of how to search for “All AIML-based Studies” and what keywords you would need to use
- Recall issues of context and consider that “triggers” are often stop words
  - “In the event that the subject is a patient with coronary artery disease but no prior family history of diabetes”
  - Consider *“no”* in TF-IDF and probability-based retrieval schemes
  - This is actually still a problem even for LLMs today
- Use TF-IDF-based IR optimized for high recall as first pass, then heavier NLP models to perform a secondary screen
  - As long as a human reviewer independently annotates a subset and agreement is high, this is considered (barely) acceptable

RICE

Use Case 3: Precision medicine as an extension to evidence-based medicine

What is the best given evidence ***fit to the patient's individual circumstances***

- Requires understanding of both the best current evidence in regards to the condition being treated, but also how *differences in patient circumstance* can change the "best" evidence

- Still an emerging field, but a common approach is to compare "similar" patients and see what worked
  - Because despite best efforts the practice of medicine is not uniform (how much "evidence" can a single clinician hold in their head?)
  - Evidence in general for a condition may not always hold true for a single patient
  - What happens if there is ***no*** suitable evidence for a given condition?

- Common approach to help with differential and treatment for rare diseases, which are very much understudied and underdiagnosed

Use Case 3: Rise of Digital Twins

What would happen to the patient if I do *x* (Simulation Only)

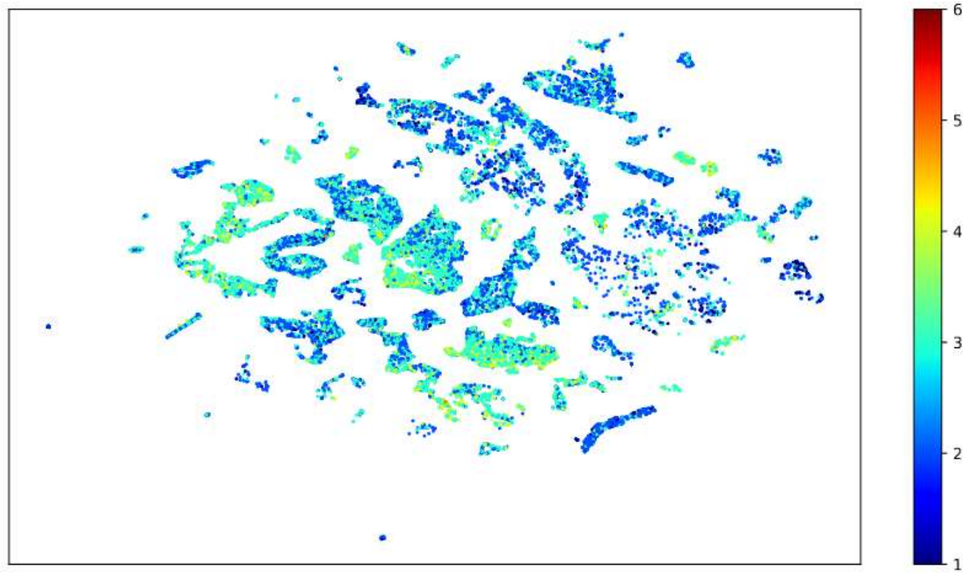
- Proposing therapies for precision medicine
- Patient trajectories and outcome comparison
- *Drug Repurposing*

- Requires both textual understanding (to find like patients) and also generation of reasoning as to why the patient is relevant/the simulated result is reasonable (for clinical evaluation)

Use Case 3: "Patients Like Me" is inherently a patient representation problem

Not every data point is relevant, but the overall representation should be similar

- This requires NLP assistance
  - 70-80% clinically relevant information in unstructured form
  - Clinician reasoning (reflecting experience) encoded in notes and sometimes departs from guidelines
    - This is why you have "#1 hospitals" such as Mayo, Cleveland Clinic, and MD Anderson
- A lot of "patient" embedding construction techniques actually borrow from NLP, even when constructed purely from structured data



Source: A visual plot of ASA assessment scores vs. patient embeddings after t-SNE dimensionality reduction. Image Adapted From: ***Wen A****, Shen F, Moon S, Liu H, Fan J. A Deep Profiling and Visualization Framework to Audit Clinical Assessment Variation. 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), Rochester, MN, USA, 2020, pp. 546-551, doi: 10.1109/CBMS49503.2020.00109.*

The Applications: Clinical Natural Language Processing supporting Healthcare AIML Translation for Clinical Practice

RICE

Use Case 1: So you want to build a risk prediction model to support clinical decision making (clinical decision support)

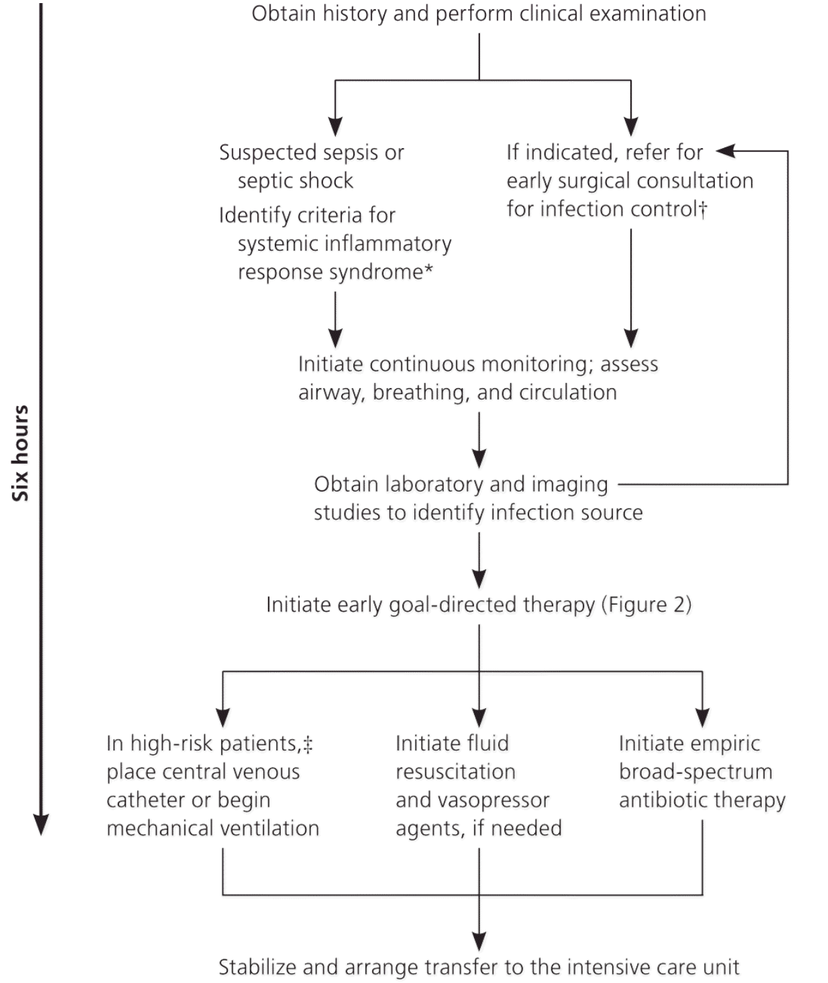
What do you need?

Research and Development:

- Data: Relevant Features, Labels: 3-12 Months
- The Model Itself: You are responsible for this

Translation/Application:

- What happens when your structured data is missing, or documentation practices change?



—Two out of four criteria must be present to identify systemic inflammatory response syndrome: fever, tachycardia, tachypnea, and leukocytosis or leukopenia.

†—Surgical intervention required for intra-abdominal perforation, obstruction, abscess, or necrotizing infection. Infected devices usually require removal.

‡—High-risk patients have systolic blood pressure < 90 mm Hg after 20 to 40 mL per kg volume challenge or lactate level > 36 mg per dL (4 mmol per L).

Management of Sepsis. Adapted from "Early Recognition and Management of Sepsis in Adults: The First Six Hours" by Robert Gauer, 2013, Am Fam Physician, 2013;88(1):44-53. https://www.aafp.org/pubs/afp/issues/2013/0701/p44.html

RICE

Use Case 1: Clinical Information Extraction

If one of your features is patients w/ fever, you need to be able to identify whether this is a yes or no!

- Requires extracting from notes
- But representations sometimes subtle
  - “Fever”, “Febrile”
  - “Temperature of *X*”
  - “Taking *x* to control symptoms”
- Oftentimes need to *impute* whether or not the feature is present based on documented clinical reasoning, rather than it explicitly

- This is all dependent on NLP!

Try it Yourself! Does this Patient have Hypertension?

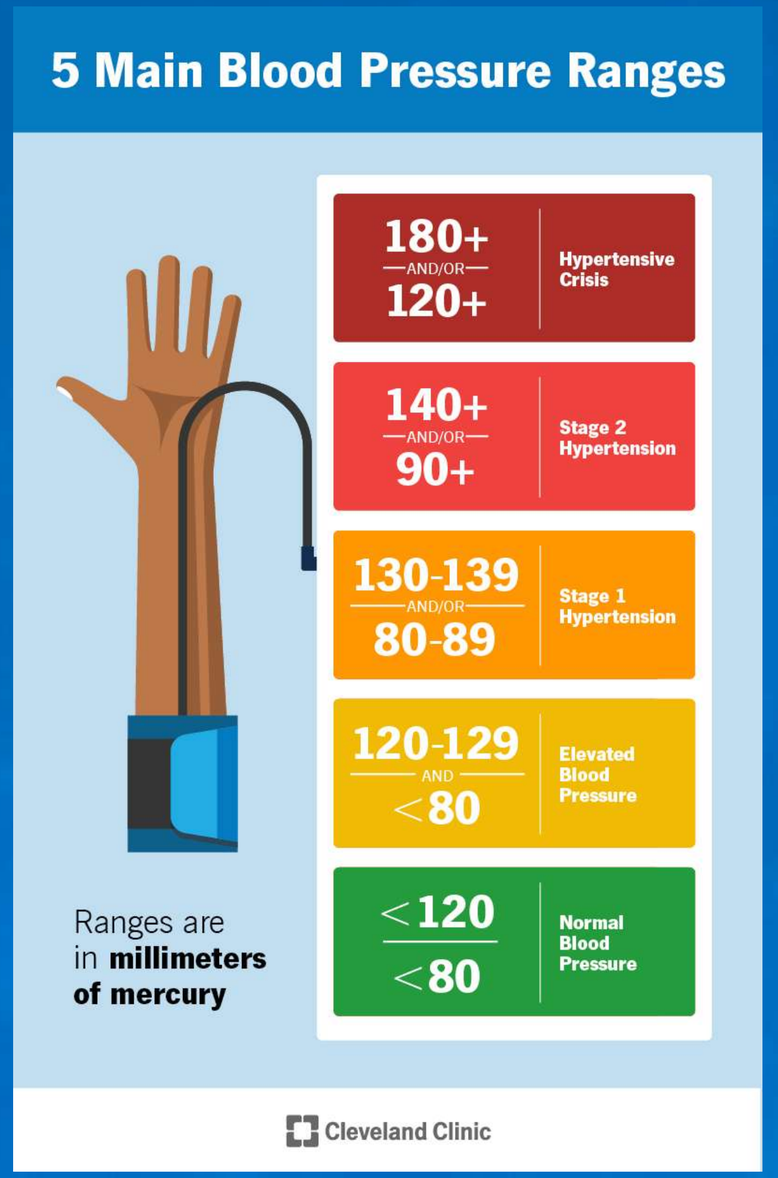
CHIEF COMPLAINT:
F/U Anaphylaxis Episode/Hypersensitivity to Unknown Allergen

PROBLEM LIST:
- Hypertension

PHYSICAL EXAMINATION:
Vital Signs: Weight: 227.2 pounds. Blood pressure: 130/79. Pulse: 80. Temperature: 97.5 degrees.
General Appearance: Not in acute distress.
Mouth: Posterior pharynx is clear.
Neck: Without adenopathy or thyromegaly.
Chest: Lungs are resonant to percussion. Auscultation reveals normal breath sounds.
Heart: Normal S1 and S2 without gallops or rubs.
Abdomen: Without masses or tenderness to palpation.
Extremities: Without edema.

MEDICATIONS:
- Loratadine 10mg / Pseudoephedrine Sulfate 240mg Extended Release Oral Tablet (Claritin-D) QD

(Note partially adapted from mtsamples, image adapted from the Cleveland Clinic)



Try it Yourself! Does this Patient have Hypertension?

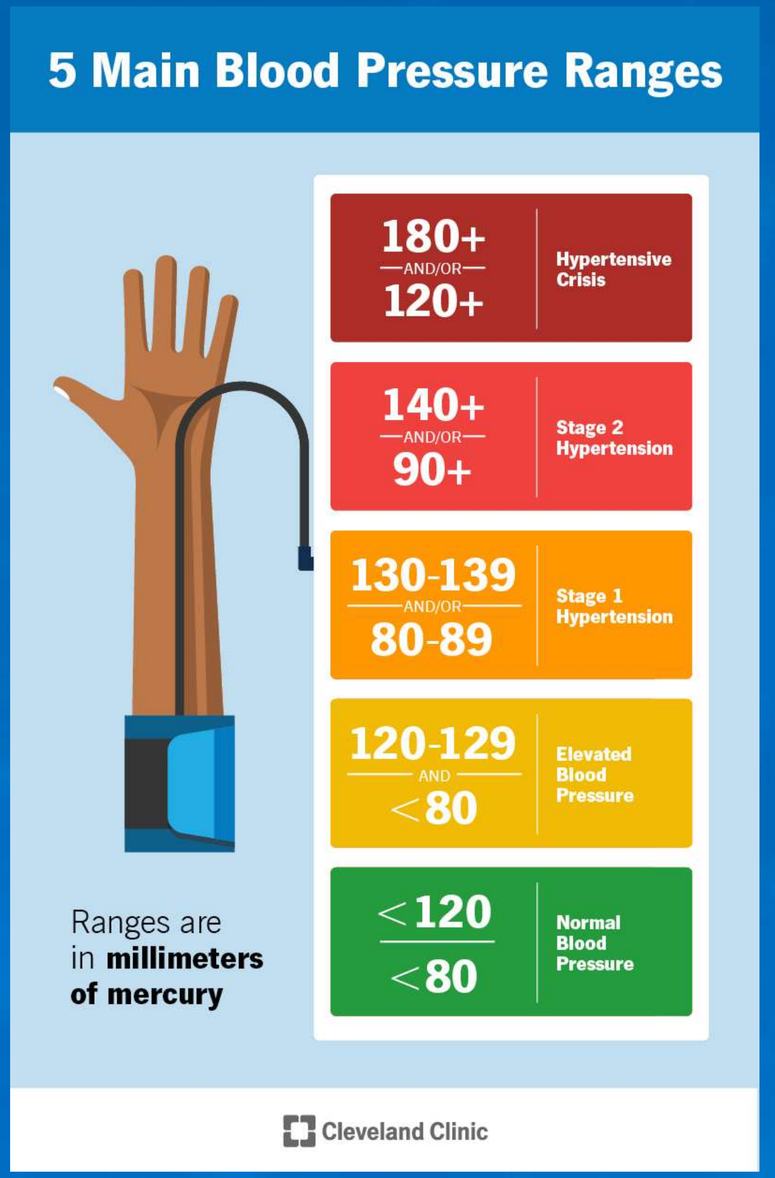
CHIEF COMPLAINT:
F/U Anaphylaxis Episode/Hypersensitivity to Unknown Allergen

PROBLEM LIST:
- Hypertension

PHYSICAL EXAMINATION:
Vital Signs: Weight: 227.2 pounds. Blood pressure: 130/79. Pulse: 80. Temperature: 97.5 degrees.
General Appearance: Not in acute distress.
Mouth: Posterior pharynx is clear.
Neck: Without adenopathy or thyromegaly.
Chest: Lungs are resonant to percussion. Auscultation reveals normal breath sounds.
Heart: Normal S1 and S2 without gallops or rubs.
Abdomen: Without masses or tenderness to palpation.
Extremities: Without edema.

MEDICATIONS:
- Loratadine 10mg / Pseudoephedrine Sulfate 240mg Extended Release Oral Tablet (Claritin-D) QD

(Note partially adapted from mtsamples, image adapted from the Cleveland Clinic)



Try it Yourself! Does this Patient have Hypertension?

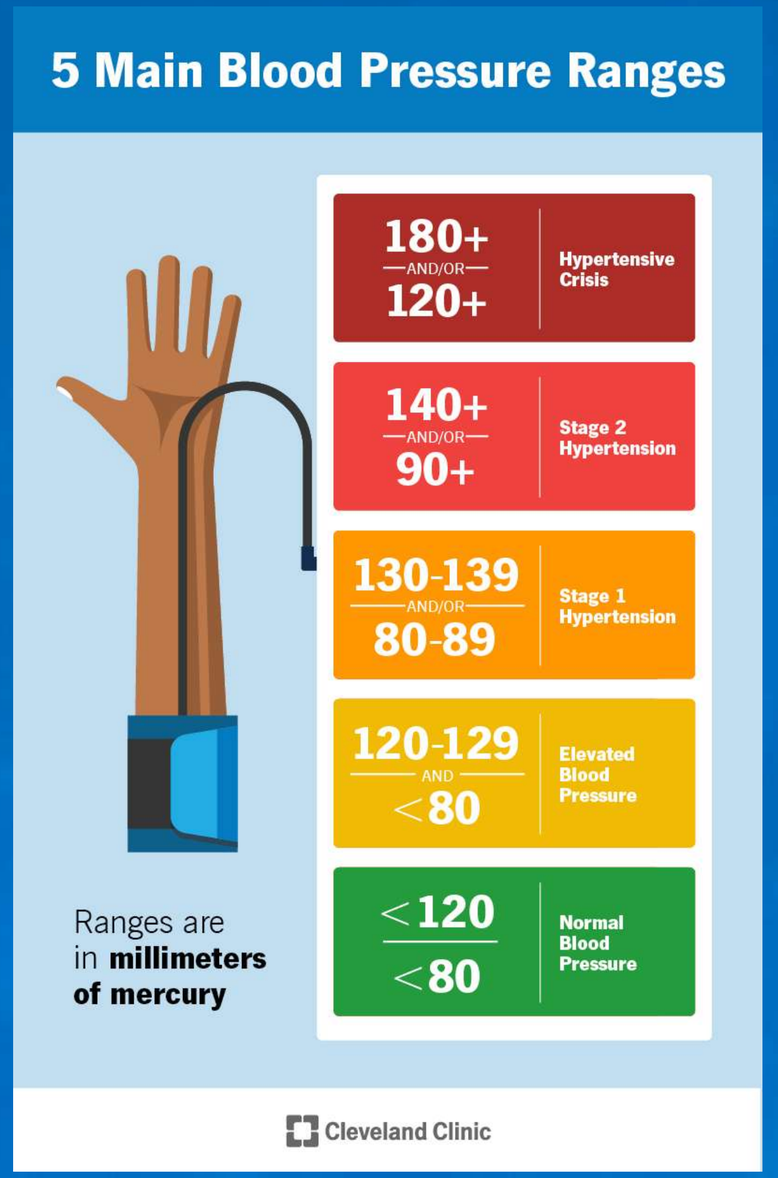
CHIEF COMPLAINT:
F/U Anaphylaxis Episode/Hypersensitivity to Unknown Allergen

PROBLEM LIST:
- Hypertension

PHYSICAL EXAMINATION:
Vital Signs: Weight: 227.2 pounds. Blood pressure: 130/79. Pulse: 80. Temperature: 97.5 degrees.
General Appearance: Not in acute distress.
Mouth: Posterior pharynx is clear.
Neck: Without adenopathy or thyromegaly.
Chest: Lungs are resonant to percussion. Auscultation reveals normal breath sounds.
Heart: Normal S1 and S2 without gallops or rubs.
Abdomen: Without masses or tenderness to palpation.
Extremities: Without edema.

MEDICATIONS:
- Loratadine 10mg / Pseudoephedrine Sulfate 240mg Extended Release Oral Tablet (Claritin-D) QD

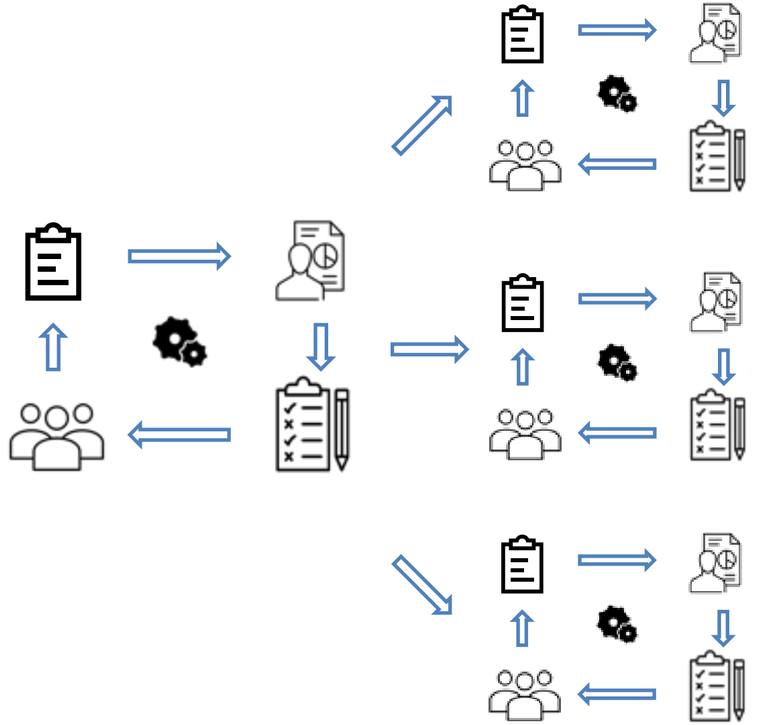
(Note partially adapted from mtsamples, image adapted from the Cleveland Clinic)



Use Case 1: Algorithm Portability has Been Major Bane in AI4Healthcare

- **Multi-institutional portability:** measures the out-of-the-box and post-refinement performance of a model when deploying it to EHRs from different institutions

- Previous studies have found between 4% to 72% f1-score degradation in various domains
  - Mental health (Cusick, JADR, 2022)
  - Cardiology (Adekkanattu, AMIA 2019)
  - Orthopedics (Han, JMIR, 2022)
- This is for the "solved" problem of Clinical Named Entity Recognition!
- Clinical linguistics differ between institutions, necessitating federated model training



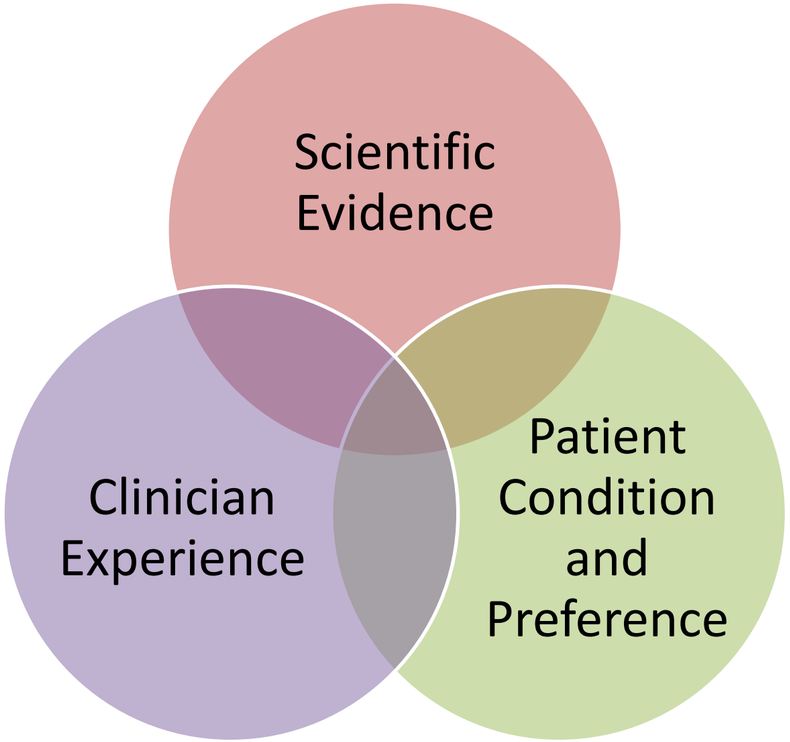
* Slide Courtesy of Sunyang Fu, UTHealth Houston

Use Case 2: "Superhuman" Clinical AIML is Fundamentally Incompatible with the Practice of Medicine

The current approach to clinical care

Evidence-Based Medicine: a paradigm of thought in clinical care where treatment decisions are made via clinicians combining what is observed about a patient alongside current best evidence in research (e.g., found through PubMed) and their own clinical experience.

- EBM has dominated clinical care practice in the past thirty years
- If known, apply knowledge, if unknown, look it up!



Use Case 2: "Superhuman" Clinical AIML is Fundamentally Incompatible with the Practice of Medicine

What enables AIML to be "superhuman"?

- Pattern recognition from data at high dimensionality
  - In particular non-linear patterns
- Less/no need for feature selection (esp. for NN-based methods)

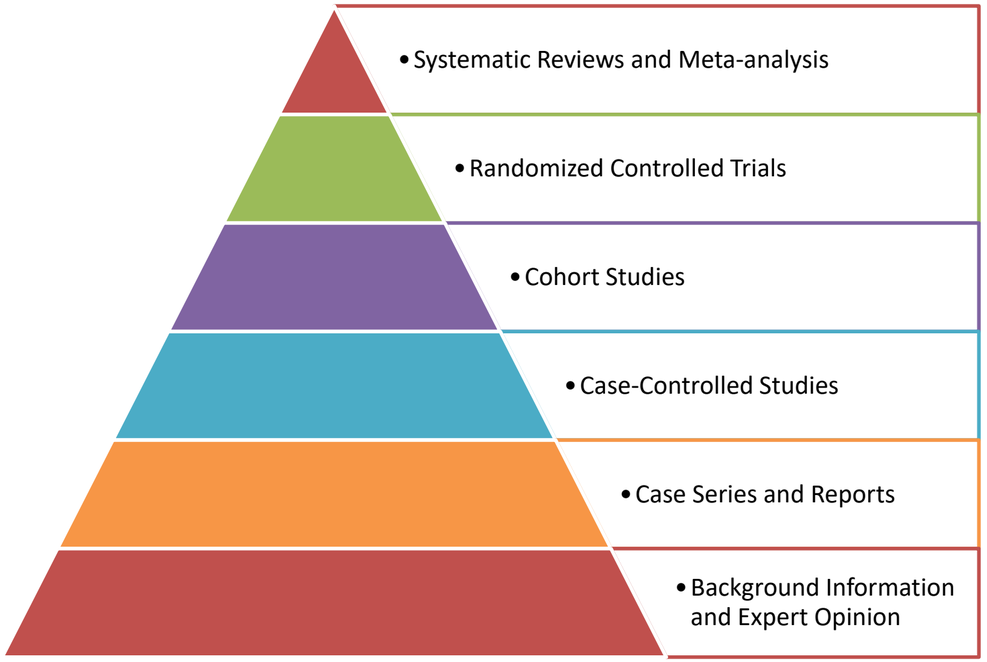
- In short – the discovery and retention of (possibly novel) associations that is consistently applied
- Inferred association to fill in "missing" information

Use Case 2: "Superhuman" Clinical AIML is Fundamentally Incompatible with the Practice of Medicine

The contradiction between superhuman AIML and evidence-based medicine

Even with perfect interpretability, insights derived from AIML methods are ***not usable*** in clinical practice without formal grounding/support in literature

- If such support existed and was consistently defined, would the AIML model be "superhuman"? (Assuming perfect application)
- Where do AIML insights fit on the pyramid of "evidence reliability"?



RICE

Use Case 2: Formalizing insights into a knowledge graph and grounding edges against PubMed at scale

The problem rephrased in a biomedical sense

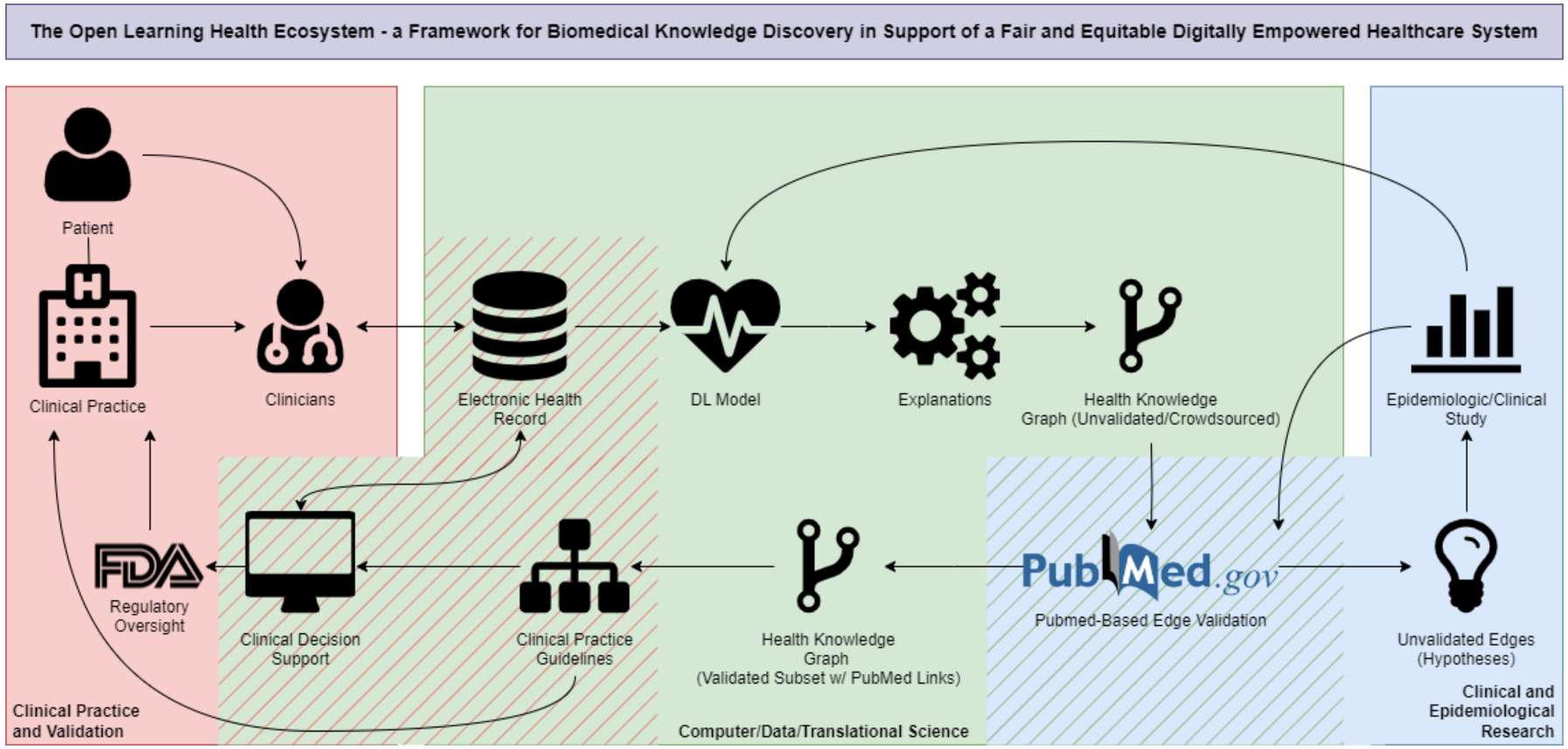
Assertion: The pathophysiology and biological mechanism of diseases follows a series of cause and effect.

- AIML "superhuman" capabilities stem from:
  - Some elements of the cause-and-effect chain being missing from human knowledge
  - AIML learning this cause-and-effect chain while being able to infer/impute missing intermediate nodes from other data
  - Superhuman recall of learned knowledge

So…. Formalize the cause-and-effect chain

- Requires mining "reasoning" from notes and biomedical literature!
- Conveniently also helps w/ some current big problems (LLM hallucination, retrieval, etc.)

Use Case 2: Formalizing insights into a knowledge graph and grounding edges against PubMed at scale

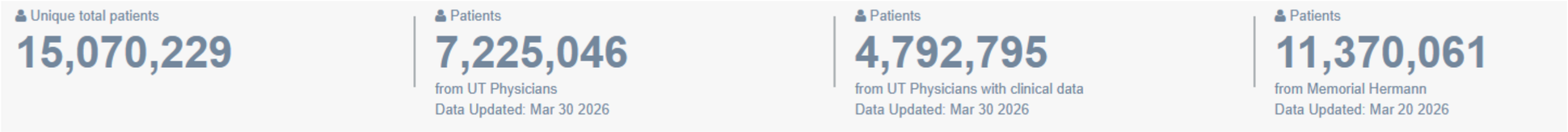


RICE

Use Case 2 Extension: Running large models on clinical data is not feasible

An individual patient can have tens to hundreds of notes

- You are dealing with patient at the bedside, do you have time to wait for an LLM to process these 100 notes (often needs to be segmented)
- Preprocessing: we have over 1 petabyte of notes, and often needs to be a comprehensive screen (e.g. for IR)
- GPUs aren't free: BERT-based deidentification of clinical text: estimated 6 months for our current clinical note dataset on 8xH100

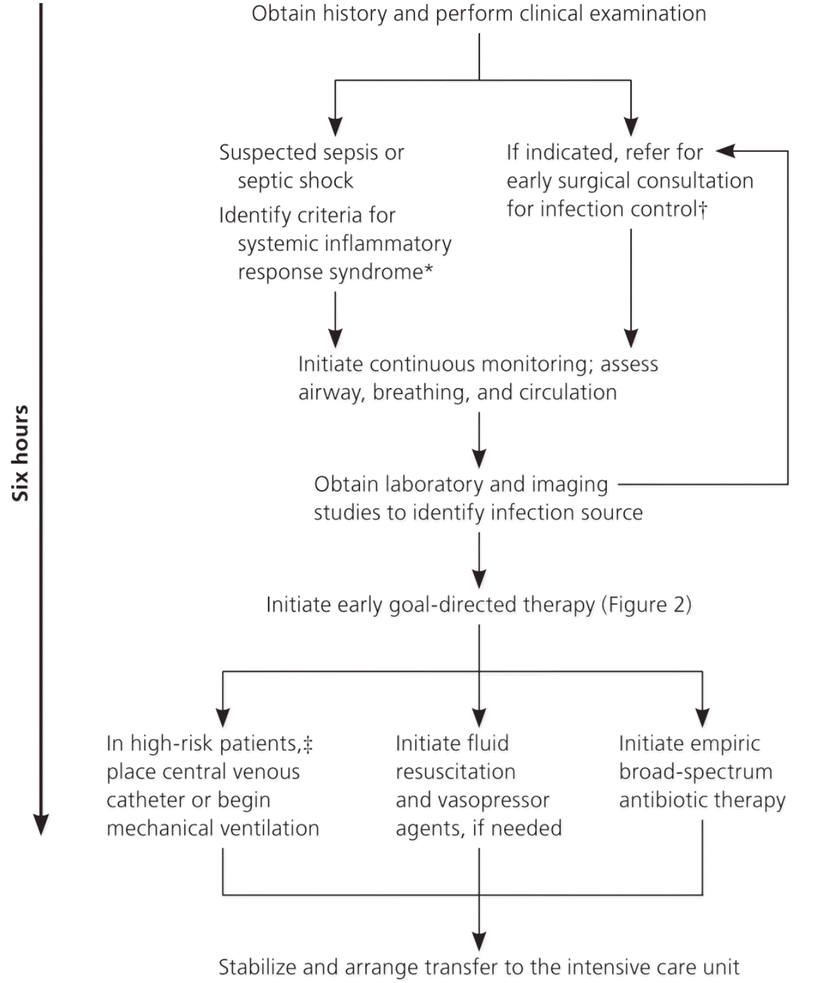


Source: https://big.uth.edu/bigarc/

Use Case 2 Extension: Running large models on clinical data is not feasible

Majority of currently deployed "AIML" models in clinical care are actually decision trees

- Use model as weak labeler, and then statistics (e.g., mutual information) to *derive* limited feature/node set and then train a decision tree instead
- Again, very dependent on data from notes *because imputation/context inferral from pure structured data is much more infeasible*
- Has benefit of facilitating algorithm portability/generalizability, and interpretability, and is much more trusted



—Two out of four criteria must be present to identify systemic inflammatory response syndrome: fever, tachycardia, tachypnea, and leukocytosis or leukopenia.

†—Surgical intervention required for intra-abdominal perforation, obstruction, abscess, or necrotizing infection. Infected devices usually require removal.

‡—High-risk patients have systolic blood pressure < 90 mm Hg after 20 to 40 mL per kg volume challenge or lactate level > 36 mg per dL (4 mmol per L).

Management of Sepsis. Adapted from "Early Recognition and Management of Sepsis in Adults: The First Six Hours" by Robert Gauer, 2013, Am Fam Physician, 2013;88(1):44-53. https://www.aafp.org/pubs/afp/issues/2013/0701/p44.html

RICE

The Frontiers: Clinical NLP in the Post-Generative Language Model Era

RICE

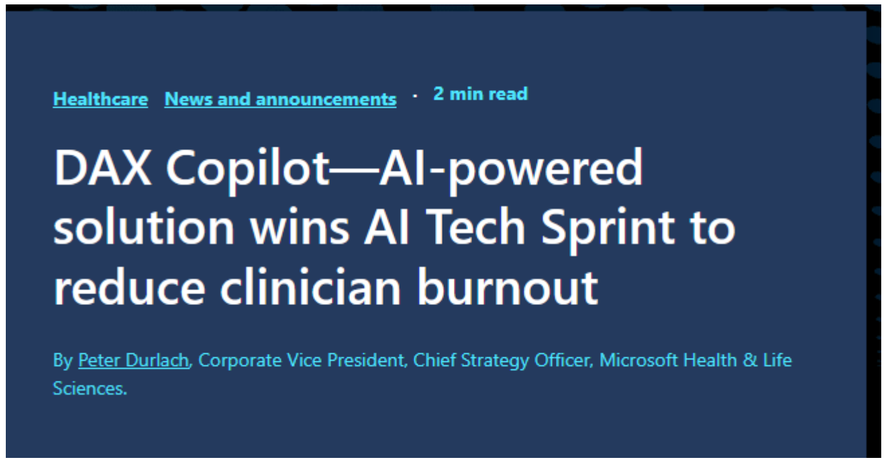
Use Case 1: Ambient Scribing

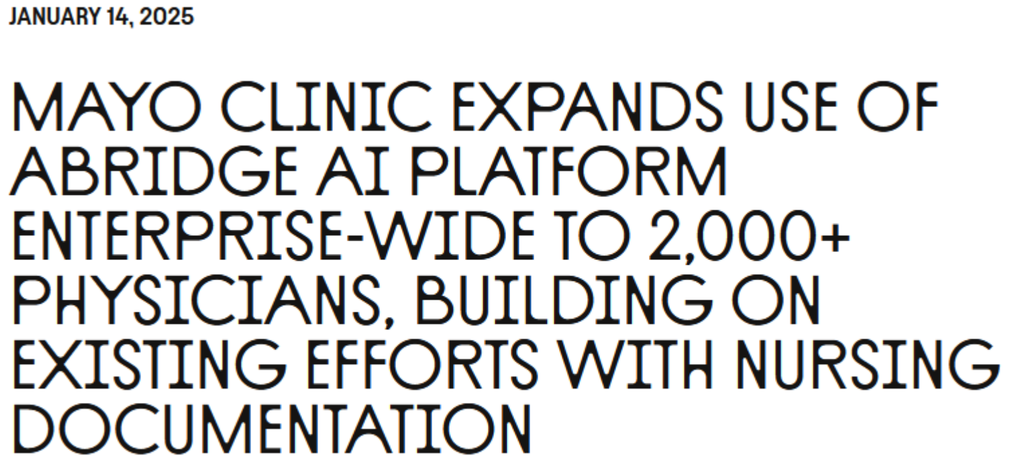
Data isn't Free: For every hour of patient care, clinicians spend two hours on documentation

- Required for regulatory and claims compliance
- A significant contributor to clinician burnout
  - An average of 1.5 to 2-hours per day of clinician "pajama-time"
- Especially in primary care and family medicine
  - The "front-lines" of clinical care with high patient volume
- Historically, dictation helps with this, but dictation is limited to transcription
- Someone also needs to ***read*** these notes
  - 10-100+ notes not unreasonable, especially for more complex patients

Use Case 1: Ambient Scribing

With Multi-modal Generative Language Models, record the patient encounter, generate summary, hands off and with information synthesis





https://www.fiercehealthcare.com/ai-and-machine-learning/epic-rolls-out-ai-charting-and-more-built-automation-clinicians-and
https://www.microsoft.com/en-us/industry/blog/healthcare/2024/06/17/dax-copilot-ai-powered-solution-wins-ai-tech-sprint-to-reduce-clinician-burnout/

Use Case 1: Ambient Scribing: Living up the Promise?

Despite being sold as such, efficacy debatable

- Very site specific, some hospitals report better, some hospitals report worse
- Dependent on *site-based customization* and ability to adapt to clinician feedback
- How do we evaluate?
  - HCI and Workflow Integration
  - Integration of prior facts/knowledge
  - Summarization Quality across Different Settings
  - Human Feedback and Model Editing
  - Patient privacy concerns for model improvement

Results

Analysis on EHR usage data from 167 physicians showed significant reductions in note-writing time, despite an increase in note length. Survey responses ($n = 65$) also indicated statistically significant improvements across multiple domains. Physicians reported reduced cognitive demand ($P = .031$) and documentation effort ($P = .014$), alongside perceptions of enhanced clinical efficiency, patient-centered care, and EHR system usability. Thematic analysis confirmed these quantitative findings and identified opportunities for improvement, including specialty-specific customization and expanded AI functionality.

Guo Y, Wang J, Hu D, Tam S, Gilman C, Chow E, Perret D, Pandita D, Zheng K. Evaluating ambient artificial intelligence documentation: effects on work efficiency, documentation burden, and patient-centered care. Journal of the American Medical Informatics Association. 2025 Oct 16:ocaf180.

RESULTS

A total of 71,487 notes were authored, of which 27,092 (38%) were generated using ambient AI. Ambient AI use had a significant reduction in work exhaustion/interpersonal disengagement (−0.44 points; 95% confidence interval [CI], −0.62 to −0.25; P<0.001), and a nonsignificant increase in professional fulfillment (+0.14 points; 95% CI, 0.004 to 0.28; P=0.04) on a five-point Likert scale. Among secondary measures, time spent on notes decreased (−0.36 hours per day; 95% CI, −0.55 to −0.17). The reduction in WoW (−0.50 hours per day; 95% CI, −0.90 to −0.09) was sensitive to exclusion of extreme values and was no longer significant after removing the top 3% of daily observations. Diagnostic billing codes improved with ambient AI use (P<0.001). Documentation quality, assessed with the PDSQI-9, demonstrated mean scores ranging from 3.97 to 4.99 across domains on a five-point scale. No drift in software performance was detected.

Afshar, M., Ryan Baumann, M., Resnik, F., Hintzke, J., Gravel Sullivan, A., Wills, G., Lemmon, K., Dambach, J., Mrotek, L.A., Quinn, M. and Abramson, K., 2025. A pragmatic randomized controlled trial of ambient artificial intelligence to improve health practitioner well-being. NEJM AI, 2(12), p.AIoa2500945.

Use Case 2: Patient Portal Messaging, Pre-Triage, and Telemedicine

- Another aspect of “documentation burden”: communication with patients
- Especially prevalent in primary care and family medicine

- Often multi-turn due to domain knowledge mismatch

- NLP can help with the following (non-exhaustive):
  - Patient routing to specialty care
  - Responses w/o full consideration of entire patient history
  - Scheduling (especially for rural access)
  - Retrospective analysis of telemedicine encounters (often dictated or through text)
  - Efficacy of telemedicine being debated (but you are charged for it!)

We're Hiring!

aw153@rice.edu

RICE