Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

How to create deployable AI in Healthcare

AIME 2025 Conference Tutorial

Tutorial Chairs

Gemma Postill,
MD/PhD Candidate
University of Toronto
No conflicts of interest to disclose.

Laura Rosella,
Professor, University of Toronto
No conflicts of interest to disclose
[Not present today]

What is T-CAIREM?

T-CAIREM: Temerty Center for Artificial Intelligence Research and Education in Medicine

Origin of T-CAIREM

The Temerty Centre for Artificial Intelligence Research and Education in Medicine (T-CAIREM)

Launched in October 2020 thanks to a generous donation from the Temerty family.

Our mission is to transform health through AI.

T-CAIREM: Structure and Leadership

Muhammad Mamdani
Director

Laura Rosella
Education Lead

Abhishek Moturu Gemma Postill
Education Student Co-Leads

Education

Infrastructure

Research

Anna Goldenberg Devin Singh
Research Co-Leads

Benjamin Haibe-Kains
Infrastructure Lead

Community

Mamatha Bhat
Engagement & Partnerships Lead

T-CAIREM Community

Over 1200 members

-**800** faculty members

-**460** students

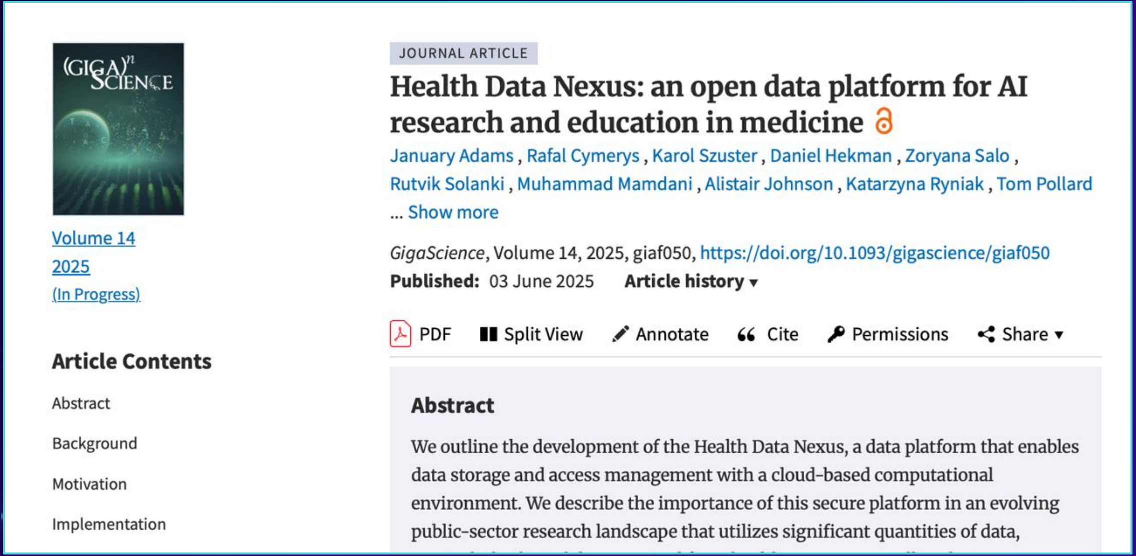
25 Canadian university partners

2 international university partners

Over 90 Canadian hospital and research institutions

T-CAIREM Health Data Nexus

One part data access platform, one part analytics environment



Volume 14
2025
(In Progress)

Article Contents

JOURNAL ARTICLE

Health Data Nexus: an open data platform for AI research and education in medicine

January Adams , Rafal Cymerys , Karol Szuster , Daniel Hekman , Zoryana Salo , Rutvik Solanki , Muhammad Mamdani , Alistair Johnson , Katarzyna Ryniak , Tom Pollard
... Show more

GigaScience, Volume 14, 2025, giaf050, https://doi.org/10.1093/gigascience/giaf050
Published: 03 June 2025 **Article history** ▾

PDF | Split View | Annotate | Cite | Permissions | Share ▾

Abstract

We outline the development of the Health Data Nexus, a data platform that enables data storage and access management with a cloud-based computational environment. We describe the importance of this secure platform in an evolving public-sector research landscape that utilizes significant quantities of data,

T-CAIREM Health Data Nexus

Datathons

Global AI in Medicine Community

Winner of the 2022 AIMed Hospital/Institution of the Year Award

T-CAIREM is a member of the Alliance of Centres of Artificial Intelligence in Medicine (ACAIM) along with more than 50 of the world's leading AI in medicine centres.

Who are you!?

What country are you from?

What is your organization and role?

Is your background clinical, technical (engineering and/or computer science), or something else?

Learning Objectives

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

1. Understand key considerations for developing an ML model that could be successfully deployed in healthcare

2. Practice applying key model development and user-interface design principles to a challenge healthcare ML challenge
3.

Table of contents

Today’s Resources

GitHub Link

Case Study & Practical Example

Check out the material that we will be using for the interactive component of of today’s session

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

AI is Rapidly Advancing Healthcare

01

Model chopping block

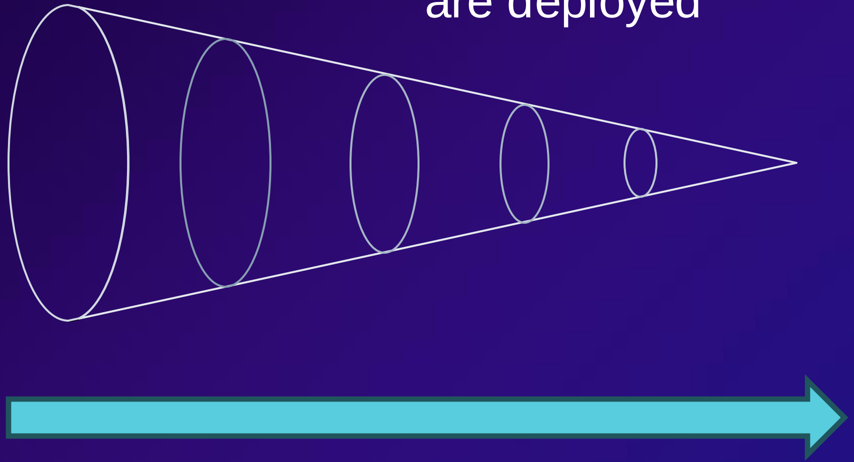
Why MOST healthcare models fail to be deployed

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

Barriers to Model Deployment

AI models that are deployed

AI Models developed for healthcare

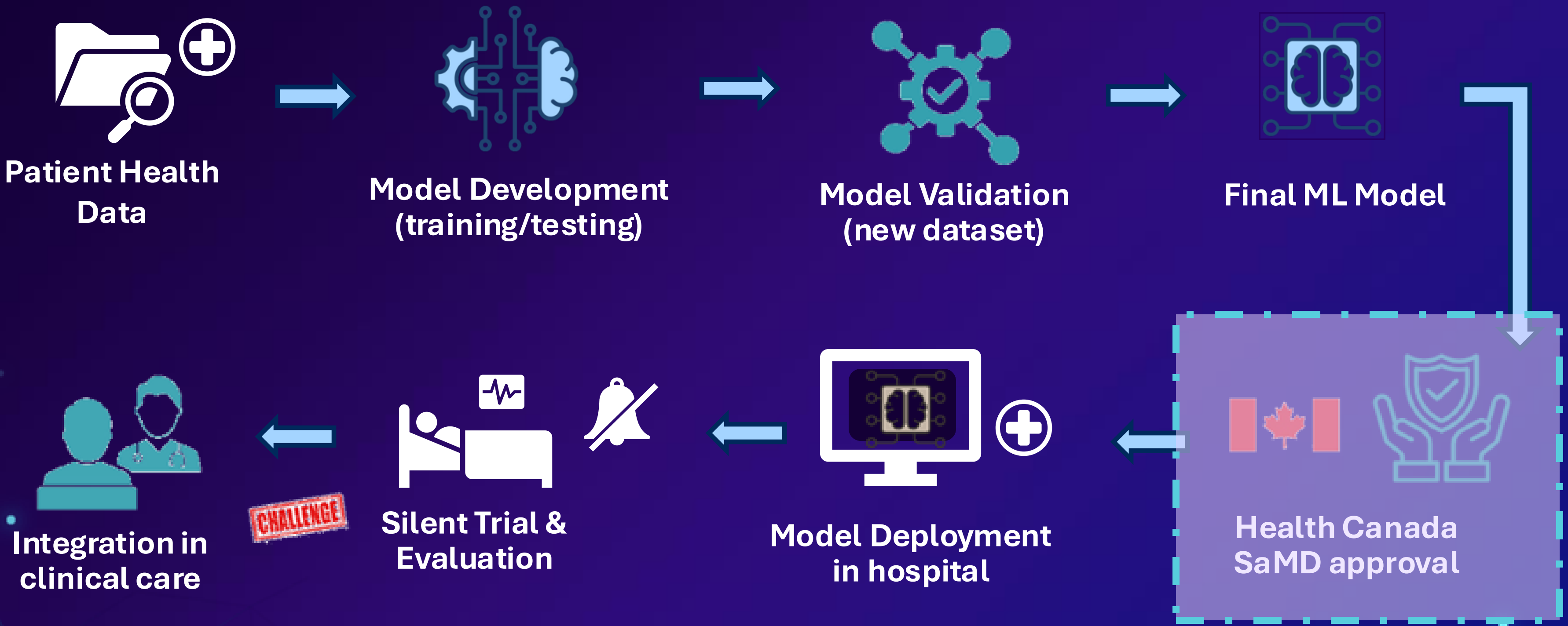


AI Models that improve** healthcare

Fewer and fewer models make it to these later stages

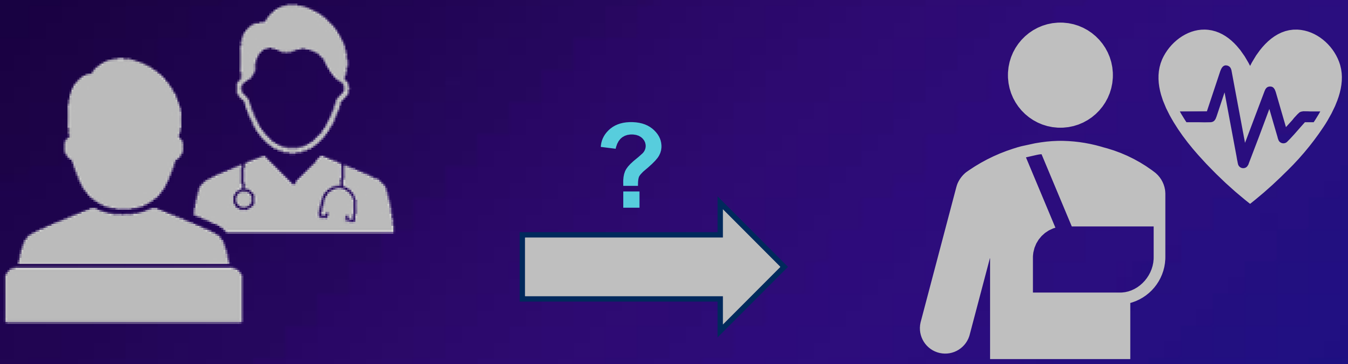
How does an ML model become clinical?

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO



How does an ML model become clinical?

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO



Integration in clinical care

Improving Outcomes?

What you think are some challenges to model deployment in healthcare?

Discussion Question

Barriers to Model Deployment

1. Misalignment with clinical priorities
2. Data quality and access
3. Lack of explainability
4. Generalizability and external validity
5. Lack of integration with clinical workflows
6. Regulatory and legal uncertainty
7. Ethical and equity concerns
8. Model maintenance and drift

Considerations for Implementing AI in Healthcare

Data	Are the data needed for the algorithm readily available and in an extractable format?
Infrastructure	Can relevant data be extracted in real time?
	Are there adequate infrastructure and computing resources available to host a cloud-based analytics and storage platform?
Interface	How will the clinical team be made aware of these predictors (i.e., is a custom dashboard integrated into the EHR required?
	How can the clinical team understand how the algorithm made this prediction (i.e., model explainability)?

Kwong et al. 2024. Nature Digital Medicine. https://doi.org/10.1038/s41746-024-01066-z

Considerations for Implementing AI in Healthcare

Continued…

End-Users	Which clinical team member(s) are most appropriate to receive the risk prediction? What is their level of trust in the model?
	What is the risk of alert fatigue, burnout, or a decrease in algorithmic compliance over time?
Clinical Context	Can an action plan be implemented based on the risk prediction?
	Are there resources available for the clinical team to action prediction?
Monitoring	How will clinical end-users or hospital leadership know whether the model continues to perform well overtime?
	What happens if performance or compliance degrades over time?

Kwong et al. 2024. Nature Digital Medicine. https://doi.org/10.1038/s41746-024-01066-z

Deployment Barriers Have Changed

Early Barriers

- Lack of favorable workflow
- Slow speed of computation
- Unrealistic expectations

Current Barriers

- Privacy and security
- Deployment accuracy
- Design/User experience
- Reliability and trust
- Ethics/Equitable benefit

Recommended Article: He et al. 2020. Deployment of Artificial Intelligence in Real-World Practice: Opportunity and Challenge

What are some solutions to these deployment barriers?

What are some solutions to Deployment Barriers?

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

- Training & Education
- Interdisciplinary design team
- Clear communication
- Iterative clinical feedback
- Regulatory compliance

When is the best time to address these barriers?

02

Case Study

Developing an AI Model to Assis with Blood Provision in Trauma

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

Traumatic Injuries

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Trauma Bay

https://www.theglobeandmail.com/canada/article-a-day-in-the-life-of-a-toronto-trauma-nurse-when-every-minute-counts/

https://unityhealth.to/2020/01/new-trauma-bay-at-st-michaels-hospital-shows-the-impact-of-simulations-on-hospital-design/

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

Blood Transfusion in Trauma

- Providing appropriate and timely transfusion therapy if critically for trauma resuscitation
- Coagulopathy is multi-factorial and heterogeneous
- Current one-size-fits-all approach to resuscitation may not be suitable for every patient
- **Rotational thromboelastometry** (ROTEM): Blood test that measures several clotting parameters to guide patient-specific transfusion needs

https://www.bbc.co.uk/news/uk-england-london-63986225

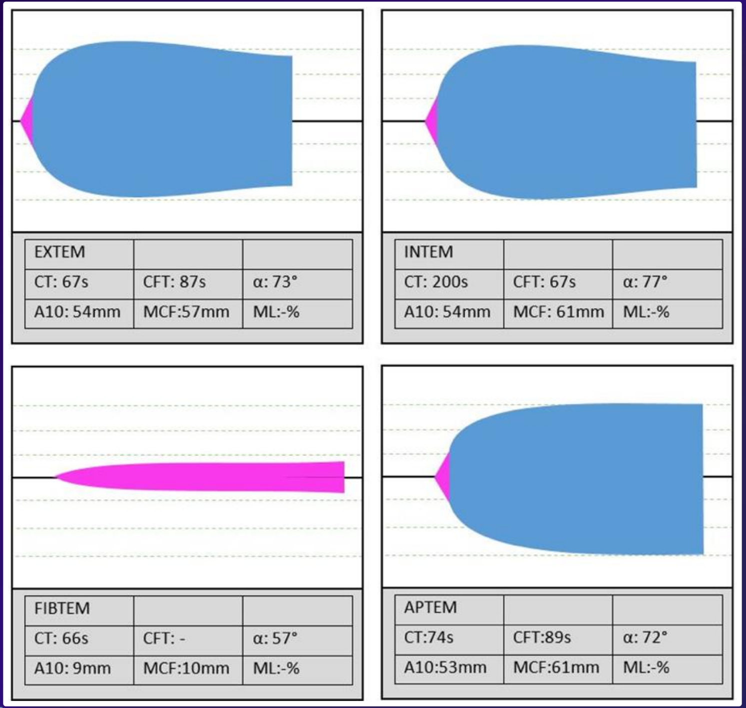
https://x.com/haemoview/status/923328278244364288/photo/2

How does ROTEM work?

- Blood sample is put into the machine
- Machine analyzes the blood
- ROTEM outputs guide transfusion of:
    1. Platelets
    2. Fibrinogen concentrate
    3. Tranexamic acid (TXA)
    4. Fresh Frozen Plasma

ROTEM Outputs are cognitively challenging to interpret

This can limit adherence clinically, particularly during acute trauma resuscitation.



https://emottawablog.com/2020/09/rotem-in-trauma-blood-is-thicker-with-wine-part-2/

ROTEM is hard to understand

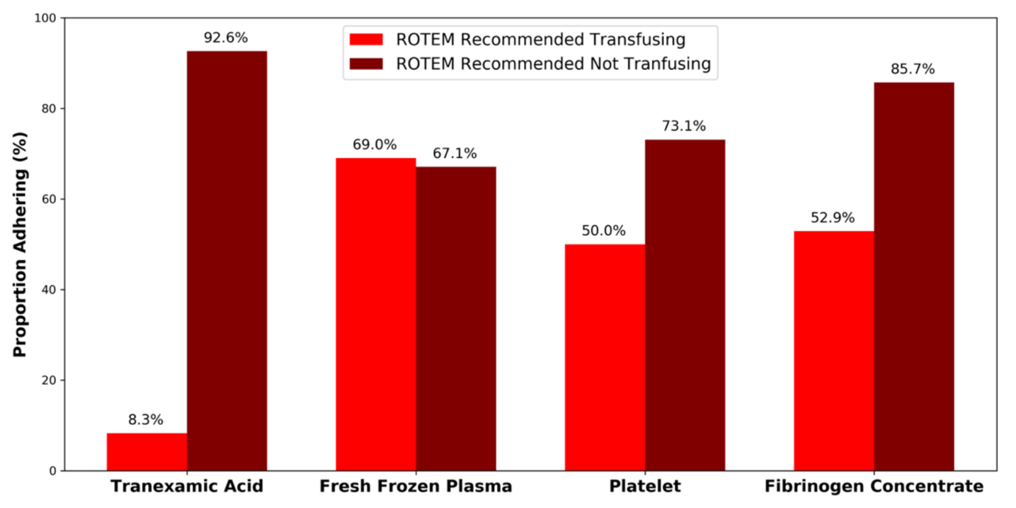


Figure 2. Adherence by blood product to ROTEM based on whether recommendation is transfusion or no transfusion (N=446).

FIRST 60
PREHOSPITAL, TRAUMA, RESUSCITATION SCIENCES

Manuscript under peer review

ROTEM is hard to understand

Especially when the patient has complex transfusion needs

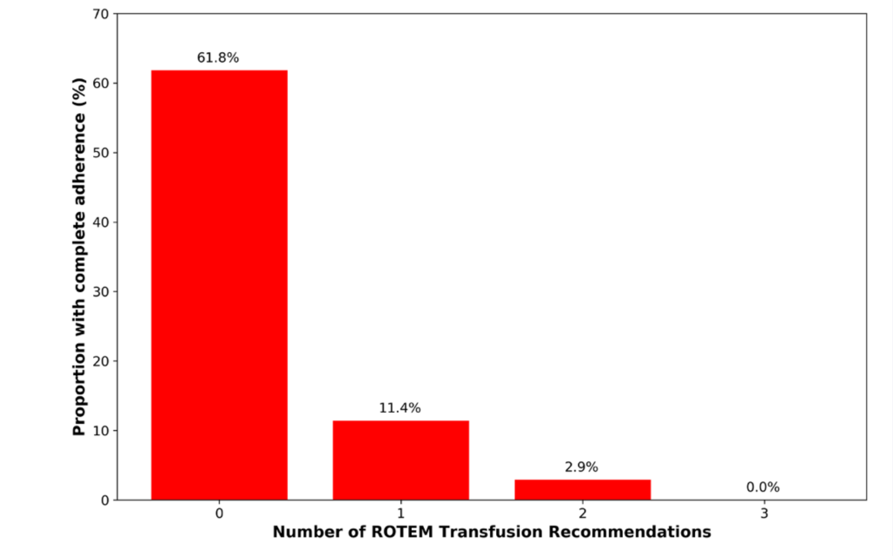


Figure 3. Complete adherence by number of ROTEM recommendations (N=446).

FIRST 60
PREHOSPITAL, TRAUMA, RESUSCITATION SCIENCES

Manuscript under peer review

03

Cultivating Success

Key inflection points for developing a deployable clinical model

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Key Inflection Points

Careful consideration of the following is needed for success

1. Clinical scenario
2. Data available
3. Prediction time window
4. Model pipeline
5. ***User interface and human-computer interactions***
6. User training and trust
7. Reporting and liability considerations
8. Post-deployment monitoring

Practical Example!

Discussed with a case study

Human-Centered Design in ML

- An approach that aims to create AI tools that are usable, ethical, and meaningful for the humans who interact with them – whether clinicians, patients, researchers, or others
- To do so, prioritizes people/s needs, values, contexts, and behaviour throughout the development and deployment of ML systems. It aims to create

Human-Centered Design in ML

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Human-Centered AI (HCAI) Framework

By Ben Shneiderman

- Encourages systems that **augment** rather than replace human intelligence.
- Three pillars: **Reliability**, **Safety**, and **Trustworthiness**.

Shneiderman, B. (2020). "Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy."

Let’s apply these considerations to our case study

#1. Clinical Scenario

Identifying a clinical scenario with an actionable, clinically relevant problem

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Identifying an Actionable Clinical Need

What makes a clinical need ”actionable”?

If the model’s output **changes what a clinician (or patient) does** – and that change has the potential to **improve care**

Identifying an Actionable Clinical Need

Ask Yourself

- Will this prediction change what the clinician does?
- Is there something specific and beneficial a clinician/patient/caregiver can do in response?
- Can that action be delivered within the time window?

What do we mean by “changing care”?

1. Improve patient outcomes
2. Improve patient experience
3. Improve efficiency (profit?)
4. Improve resource allocation
5. Improve physician experience

And more!?

Examples of (Non-) Actionable Problem

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

Actionable or not?

Predicting sepsis onset in hospitalized patients	Actionable Early detection can prompt immediate evaluation, labs, antibiotics, and fluids

Examples of (Non-) Actionable Problems

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Actionable or not?

Inferring patient's emotions from EHR text	**Not Actionable** Hard to verify or act on in a standardized, safe, or equitable way

Examples of (Non-) Actionable Problems

Actionable or not?

Predicting diagnosis in a neurodegenerative disorder with no treatment

Not Actionable

No intervention to offer, may not benefit patient to know earlier (unless there are patient supports that we can provide)

Examples of (Non-) Actionable Problem

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Actionable or not?

Classifying diabetic retinopathy from retinal images

Actionable

Can trigger referrals to ophthalmology or expedite treatment

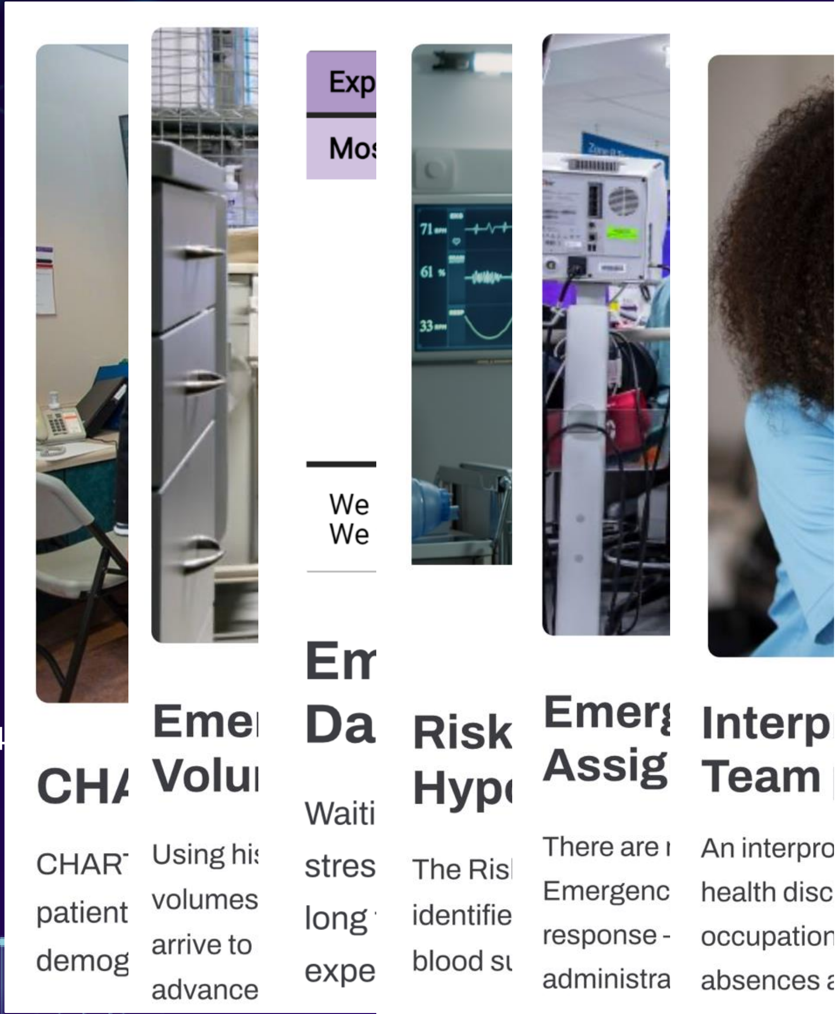
Examples of (Non-) Actionable Problem

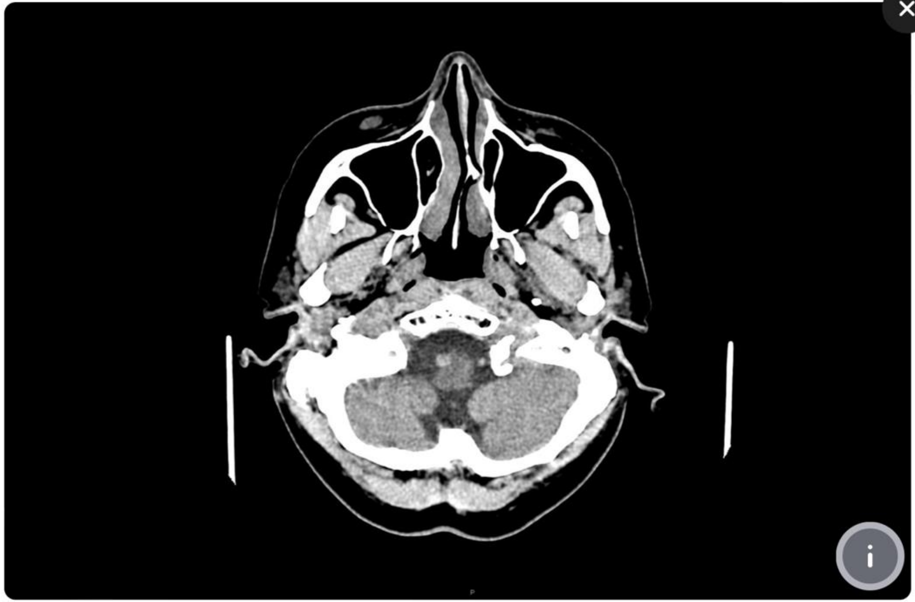
Actionable or not?

Identifying patients at high risk of readmission

Actionable

Enables transitional care planning, early follow-up, or case management





Brain Bleed Detection tool

This AI solution looks at the CT scans for St. Michael's Emergency Department patients who come in with a head injury and predicts whether the patient has an intracranial hemorrhage – also known as a brain bleed. The predictions are flagged to care providers so they can assess and give faster diagnoses and treatment.

Decision-Support vs. Decision-Making

Clarify whether the goal is to support vs. replace clinicians

Feature	Decision-Support	Decision-Making
Role	Assist clinicians	Replace or automate human decision
Human oversight	Always present	May be minimal or absent
Risk level	Lower	Higher
Regulatory oversight	Lower bar	Higher bar

Four Types of Intelligence in AI

Non-Adaptive Machines

No-Human/Little Intervention

Human Intervention

	No-Human/Little Intervention	Human Intervention
Non-Adaptive Machines	Automation Intelligence	Assisted Intelligence
Adaptive Machines (learning)	Autonomous Intelligence	Augmented Intelligence

Adaptive Machines (learning)

Vipul Gupta – https://www.linkedin.com/pulse/four-types-intelligence-ai-vipul-gupta

Applying this to our case study

#1. Clinical Scenario

Is there a clinical problem that can be improved by AI? What is it?

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

#2. Data Available

Determining what data is available to train and run the model

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

What Data Exists?

Data available to train a model might not be the same as data available clinically

Sources can include

- Hospital discharge records
- Physician billing claims
- Emergency department visits
- Prescription drug databases
- Provincial or national insurance records
- Disease registries – designed for quality improvement or research

Health Data Quality

- One of the most critical success factors
- Poor quality can introduce bias and reduce generalizability (impact on performance may not be seen until deployment)
- Quality of development and deployment data likely different

Health Data Quality

An example of the importance of label quality – multimorbidity

Science

Current Issue | First release papers | Archive | About | Submit manuscript

HOME > SCIENCE > VOL. 366, NO. 6464 > DISSECTING RACIAL BIAS IN AN ALGORITHM USED TO MANAGE THE HEALTH OF POPULATIONS

RESEARCH ARTICLE

Dissecting racial bias in an algorithm used to manage the health of populations

ZIAD OBERMEYER, BRIAN POWERS, CHRISTINE VOGELI, AND SENDHIL MULLAINATHAN   Authors Info & Affiliations

Health Data Quality

An example of the importance of label quality – multimorbidity

Cost of previous healthcare	≠	Need for future heatlhcare

Health Data Quality

An example of the importance of label quality – asthma diagnosis

Health Data Quality Considerations

- **Completeness** – what is missing? Is missingness reflective of normal practice? Could missingness be a valuable modelling signal
- **Accuracy** – includes of label (ground truth), feature values, feature/label timestamps
- **Data leakage** – any features affected by the label?
- **Standardization** – are units or diagnostic codes harmonized? Have any of these changed?
- **Representativeness** – is the data represented of the people and medical conditions you want to capture?
- **Bias in Source data** – could sociocultural biases be reflected in how the data is collected / compiled

Applying this to our case study

#2. Data Available

Is there data available?
Is usable data naturally created in clinical workflow?

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Training a Model on Blood Products Administered...

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

#3. Prediction Time Window

Timeframe of clinical problem and timepoint(s) at which prediction is relevant

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

3. Prediction Time Window

Importance of defining the prediction horizon

Data → **Prediction** → **Clinical Action**

Considerations include:

- Clinical workflows and decision timelines
- Intervention effectiveness windows
- Time constraints in real-world settings

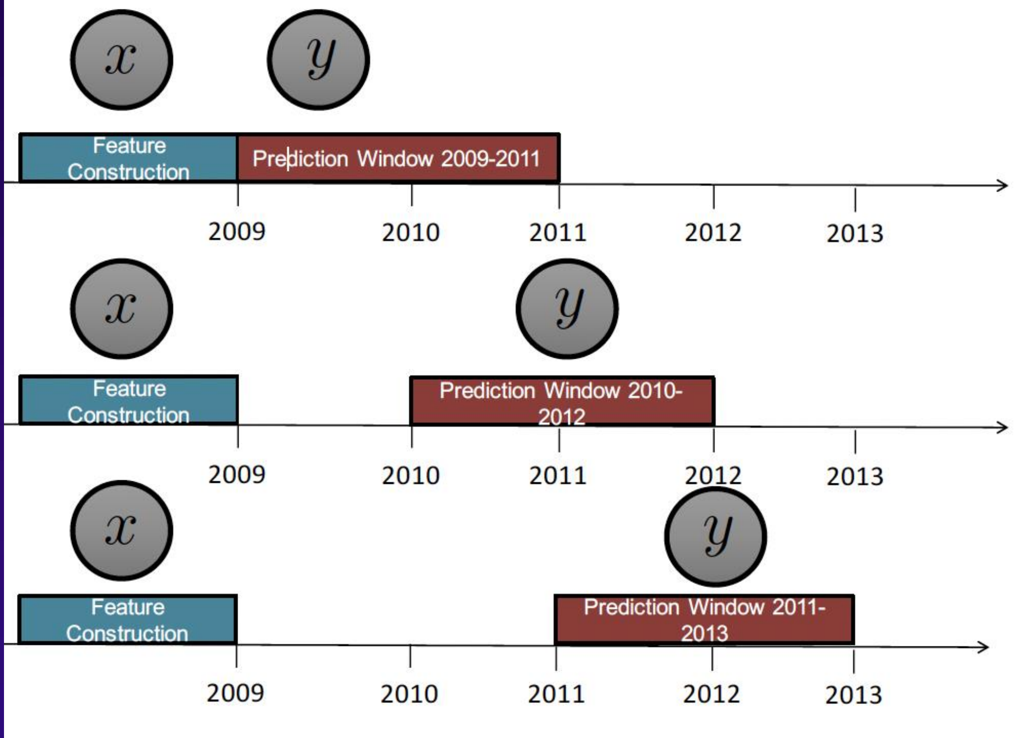
Prediction Time Window

Which of the following is the best prediction time window for a healthcare ML model?

Option A

Option B

Option C



Applying this to our case study

#3. Prediction Time Window

When is the information needed for the clinical decision?
Is there time for an actionable change?

Prediction Time Window for ROTEM

- Golden hour of trauma
- Blood is analyzed as soon as possible
- ROTEM analysis takes 5-15min

Are you predicting what ROTEM will say?

- ROTEM output is more relevant at time of analysis (blood clotting values will change with administration of blood products)\
- ROTEM results would NOT be used to guide transfusion >24 hours after analysis

Are you interpreting the ROTEM output?

#4. Model Pipeline

Developing the *right* model

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

Model Input and Output

Considerations include:

- Is input image or tabular based? Other format?
- Is output categorical or numerical (regression vs. classification)?

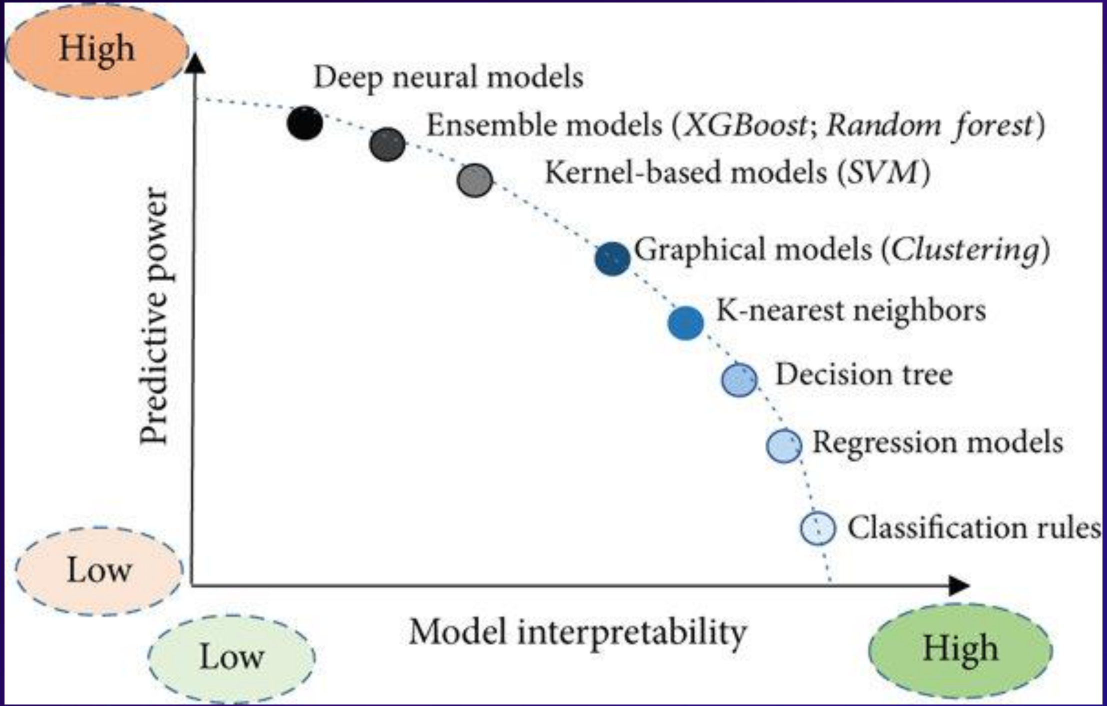
What type of Model are you Developing?

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

	Tool user	
Output	Healthcare professional facing, generative AI model	Patient professional facing, generative AI model
	Healthcare professional facing, non-generative AI model	Patient professional facing, non-generative AI model

Predictive Power vs. Interpretability

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO



https://www.researchgate.net/figure/Predictive-power-vs-interpretability-trade-off_fig4_352986044

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

Computational Power vs. Simplicity

medRχiv

THE PREPRINT SERVER FOR HEALTH SCIENCES

Follow this preprint

DynaMELD: A Dynamic Model of End-Stage Liver Disease for Equitable Prioritization

Michael J. Cooper, Xiang Gao, Xun Zhao, Dariia Khoroshchuk, Yingke Wang, Amirhossein Azhie, Maryam Naghibzadeh, Sandra Holdsworth, Jed Adam Gross, Michael Brudno, Jordan J. Feld, Elmar Jaeckel, Gideon Hirschfield, Rahul G. Krishnan, Mamatha Bhat

doi: https://doi.org/10.1101/2024.11.19.24316852

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should *not* be used to guide clinical practice.

Abstract | Full Text | Info/History | Metrics | Preview PDF

https://www.medrxiv.org/content/10.1101/2024.11.19.24316852v1

Applying this to our case study

#4. Model Pipeline

Are there viable ML model architectures that can predict this output?

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Can ML interpret ROTEM?

Review | Open access | Publishe

Artificial intellige
hemorrhagic trau

Henry T. Peng, M. Musaab Sidc

Andrew Beckett

Military Medical Research **10**, Art

A Commentary to this article w

INDEPENDENT SUBMISSIONS

Predicting blood transfusion following traumatic injury using machine learning models: A systematic review and narrative synthesis

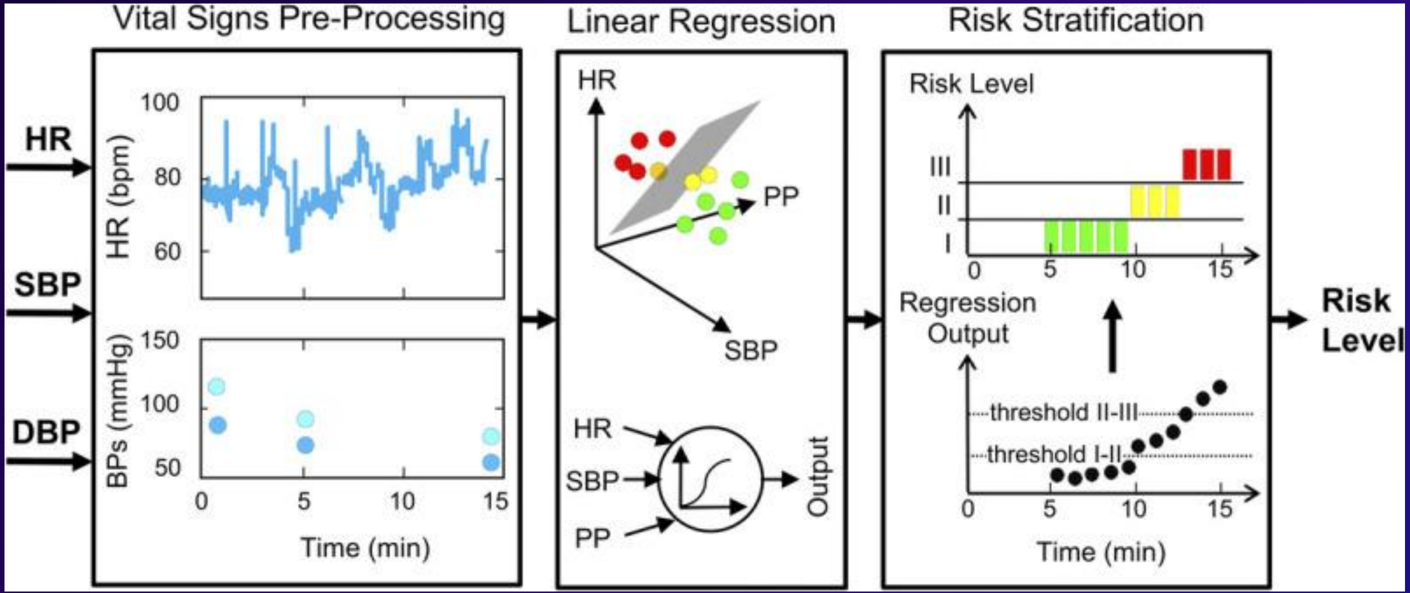
Oakley, William MSc; Tandle, Sankalp MBChB; Perkins, Zane PhD, FRCS; Marsden, Max PhD, FRCS

Author Information

Journal of Trauma and Acute Care Surgery 97(4):p 651-659, October 2024. | ***DOI:*** 10.1097/TA.0000000000004385

BUY   SDC   INFOGRAPHIC

Good performance, but deployment?



APPRAISE-HRI Model

Stallings et al. 2023. *Shock.*

What Type of Model would you Develop?

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Healthcare professional facing, generative AI model	Patient professional facing, generative AI model
Healthcare professional facing, non-generative AI model	Patient professional facing, non-generative AI model

Most has been here…

#5. User Interface

& Human-Computer Interactions

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

User Interfaces (UI)

- **UI**: the point of interaction between a human and computer system.
- UI includes all the visual elements and processes that allow users to communicate with and operate a system/model,
- The UI can take many forms – apps, websites, ML tool in a clinical setting
- Key consideration in developing a UI is who will be using it and what they will need to do with the insight

Key Components of UI

- **Input controls** – Buttons, forms, sliders, checkbox (automatic input?)
- **Navigation** – Menus, tabs, breadcrumbs
- **Information displays** – Graphs, tables, alerts
- **Feedback mechanisms**…

Key UI Design Pitfalls

- Too much data, too little context
- Overloading alerts
- Ignoring user role differences (nurse ≠ MD ≠ admin)

Great UI can make or break adoption in Healthcare

UX vs. UI

UI: User Interface

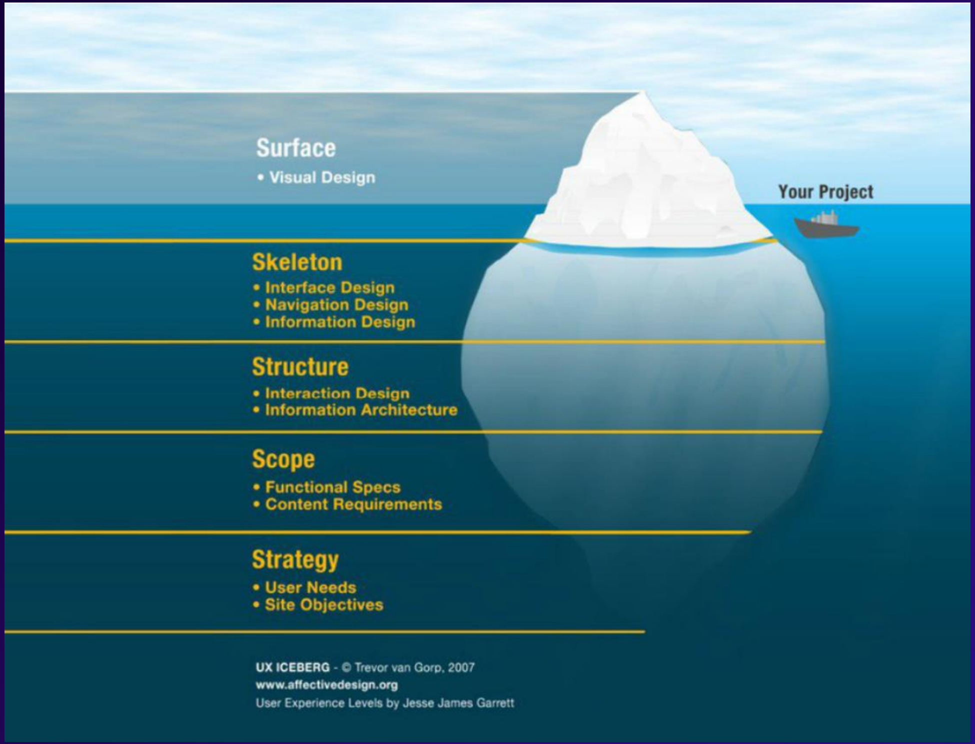
The look and feel of the product's interface – the visual design and layout of elements like buttons, colors, typography, spacing, and interactive components.

UX: User Experience

The overall interaction with a product or system – includes how they feel, how easy it is to accomplish tasks, and whether their needs are met efficiently and enjoyably

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Levels of the UX Design Process

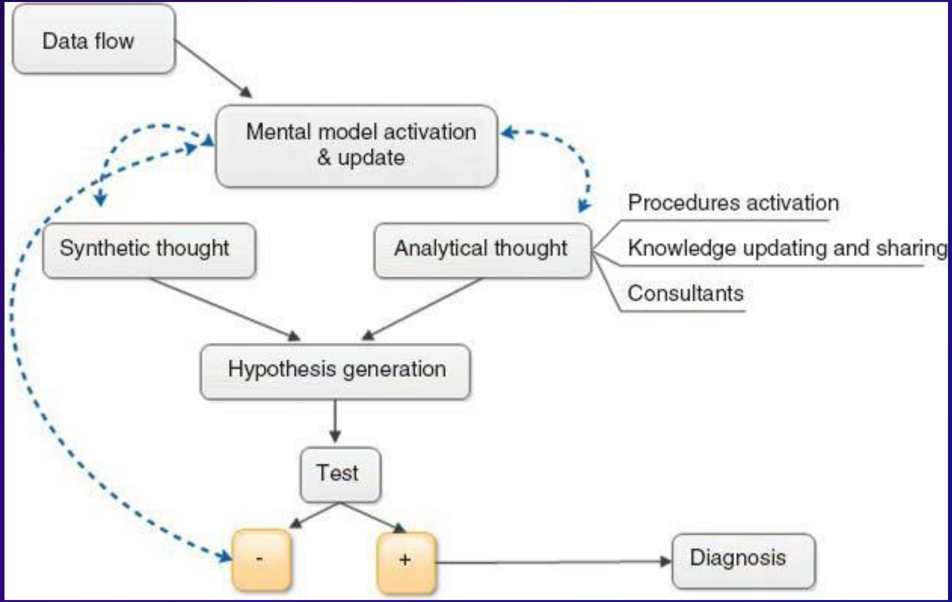


https://www.scriptutex.pt/2019/01/10/5-levels-of-ux-design-process-2/

Starts at the bottom

Recognizing the Clinician's Mental Model

- **Mental Model**: way of conceptualizing how they interpret data
- Important re: how they expect a tool to behave



https://www.researchgate.net/figure/Mental-model-is-the-physicians-cognitive-structu that-incorporates-and-gives-sense-to_fig1_264676002

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

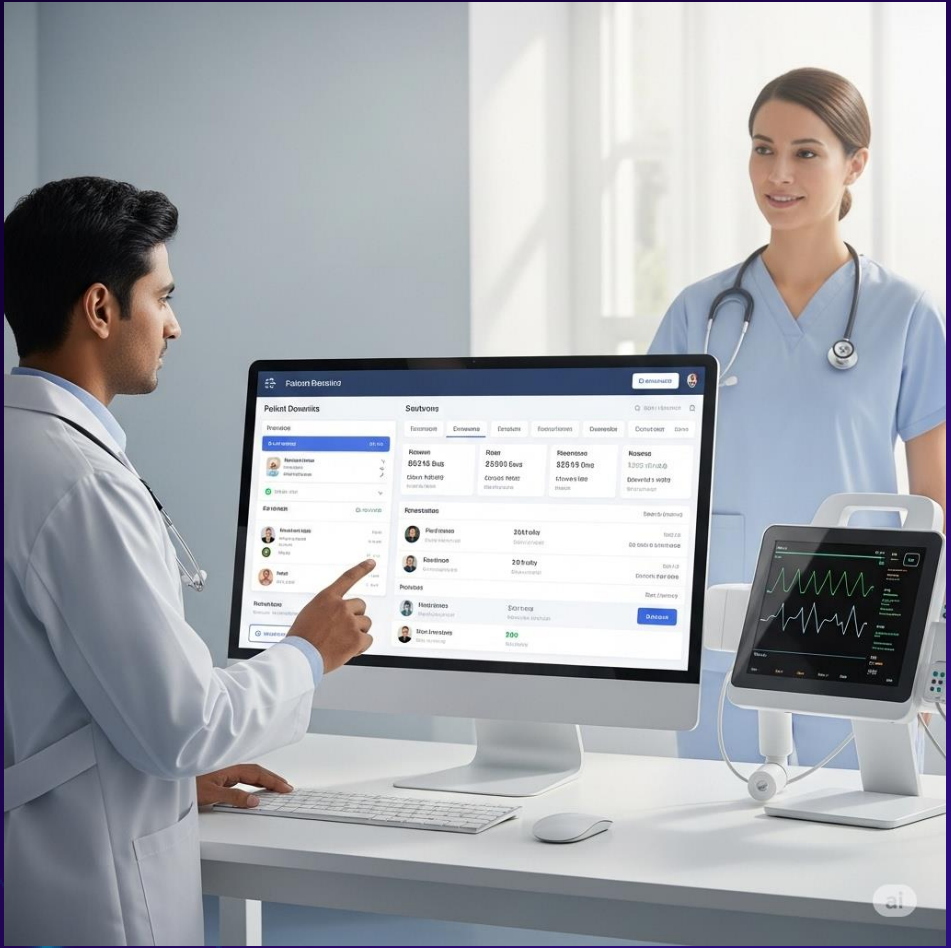


Image Generated by Gemini

EHR Integration is Key

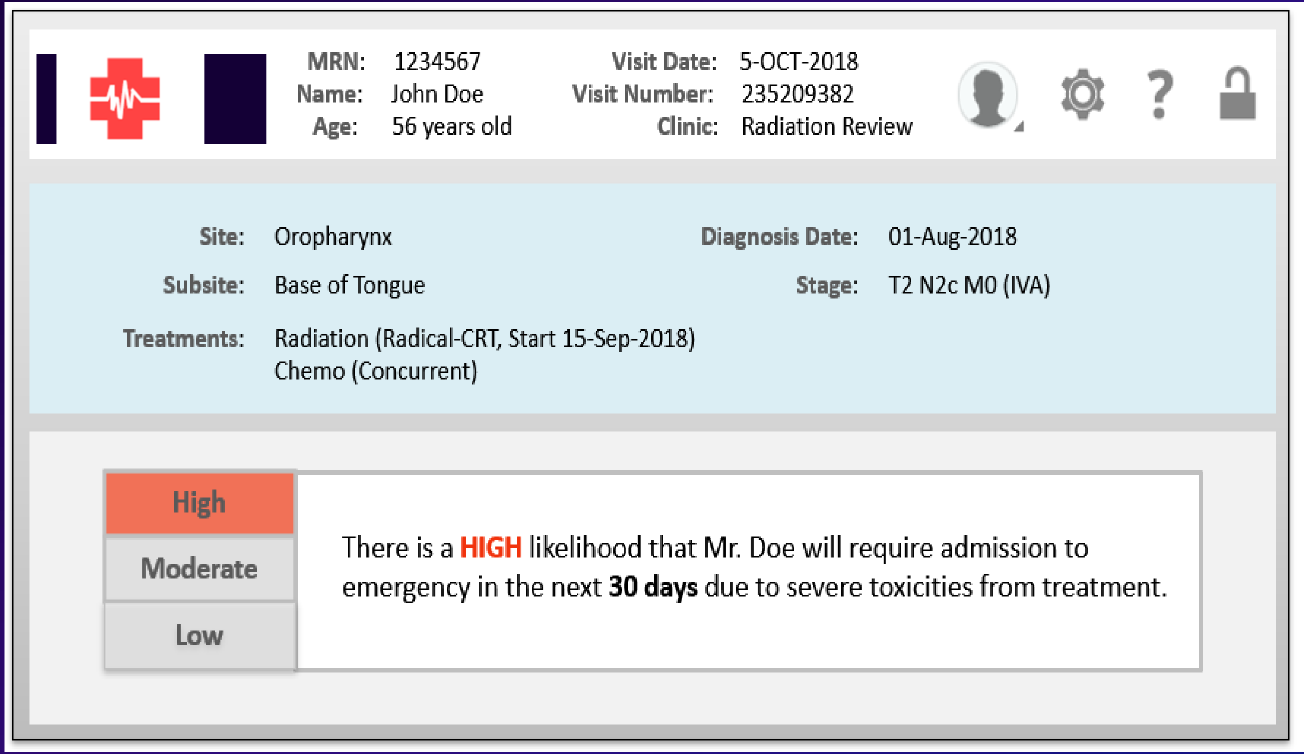
- Stand alone dashboards often fail
- Workflow-aware design is key

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Anatomy of a Clinician-Facing AI Tool

What is good about this interface?

What could be improved?



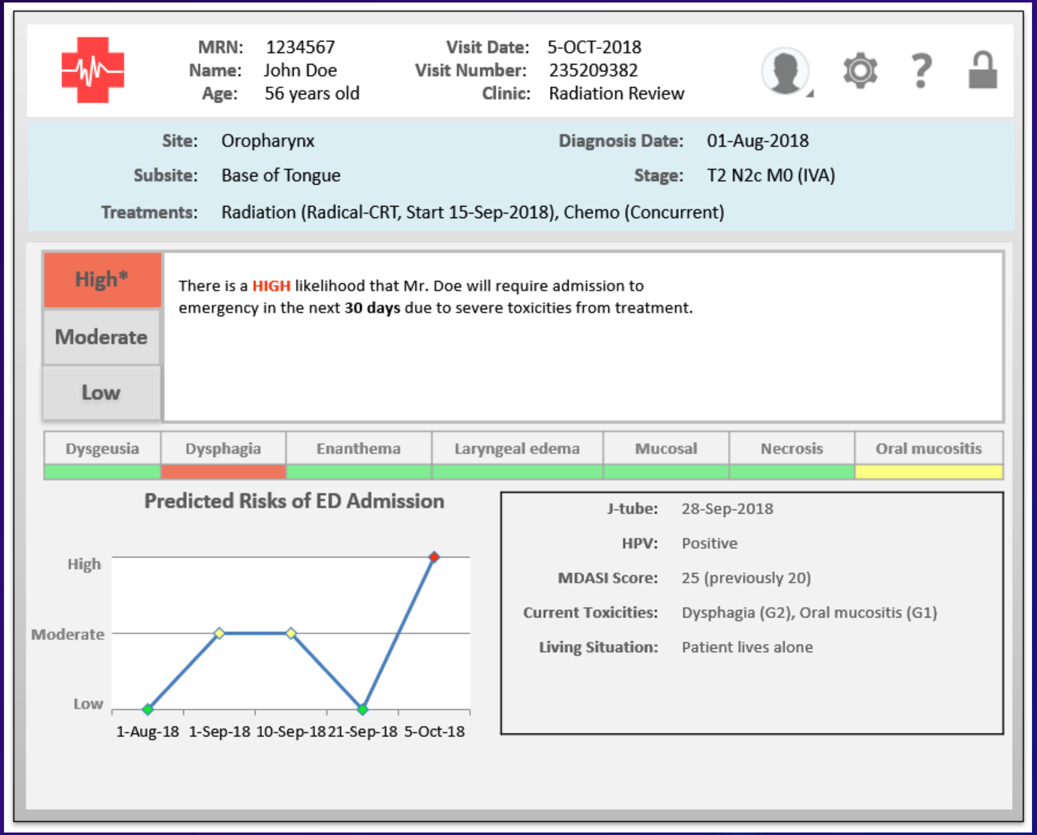
Modified Material From: *AI Development and Implementation in Health Care **(LECUTRE:** End user interfacing with AI, Tran Truong)*

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Anatomy of a Clinician-Facing AI Tool

What is good about this interface?

What could be improved?



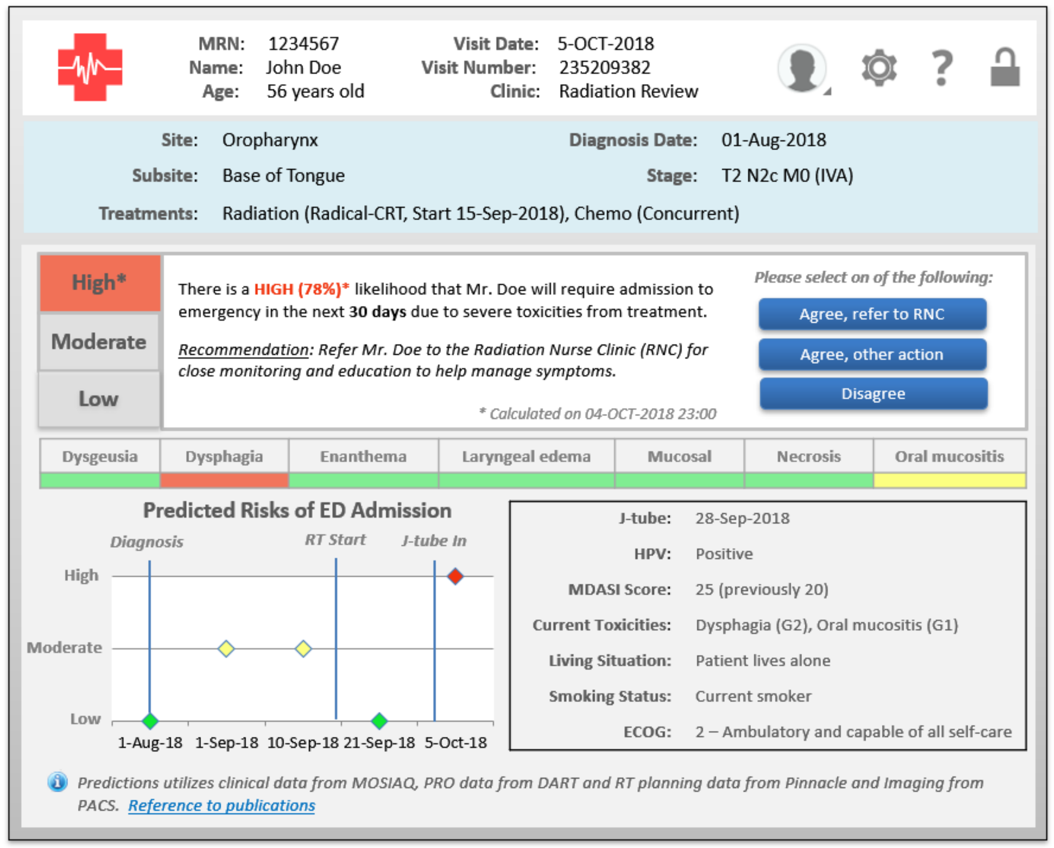
__Modified Material From:__ Gillbank P, et al. Designing for Physician Trust: Toward a Machine Learning Decision Aid for Radiation Toxicity Risk: Ergonomics in Design., 2019. doi:10.1177/1064804619896172

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Anatomy of a Clinician-Facing AI Tool

What is good about this interface?

What could be improved?



Modified Material From: *Gillbank P, et al. Designing for Physician Trust: Toward a Machine Learning Decision Aid for Radiation Toxicity Risk: Ergonomics in Design., 2019. doi:10.1177/1064804619896172*

What else could be on a healthcare ML model's user interface?

Discussion Question

Consider Different UI Needs for:

Healthcare professional facing, generative AI model	Patient professional facing, generative AI model
Healthcare professional facing, non-generative AI model	Patient professional facing, non-generative AI model

Labeling & Terminology on UI

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

- Use clinical language (“readmission risk” vs. “output=0.78”)
- Use units that are familiar (%, days, mg/dL)
- Avoid ML jargon & include definitions where needed
- Prevent over automation… “AI assist” vs. “AI decide”

Conveying Uncertainty in Model Output

Methods:

- Confidence intervals
- Wording (e.g., “likely”, “possible”)

Discussion Question:

How much uncertainty is too much uncertainty?

Visual Design Basics

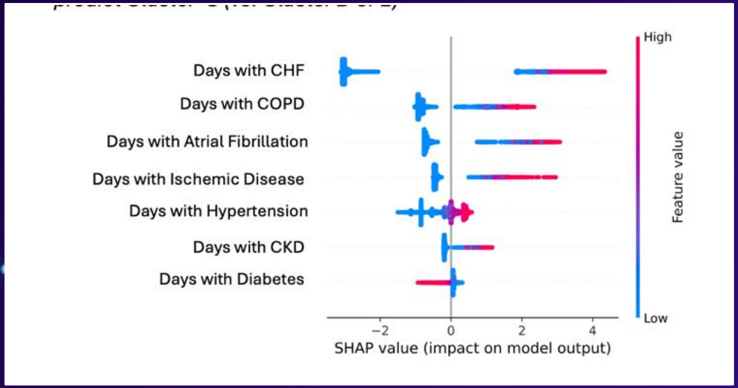
- Visualize output (e.g., probability sliders, traffic lights)
- Hierarchy design structure (e.g., bolding, grouping)
- Appropriate color schemes (e.g., red/green caution)
- Effective use of white space

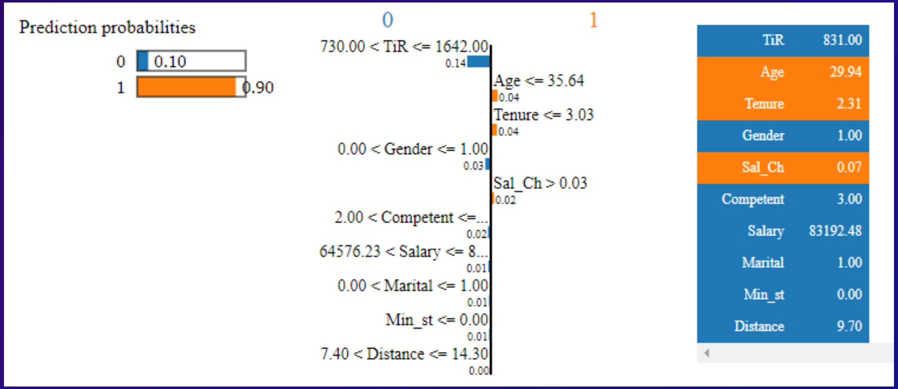
Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Including Explainable Elements:

Common approaches for model explainability:

- SHAP: Feature importance visualization
- LIME: Local surrogate models
- Counterfactuals “What if?” Scenarios





https://github.com/marcotcr/lime/issues/246

Are these methods understandable to clinician users?

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Interactivity & Customization

Clinically Relevant Examples:

- Filters
- Time windows
- Patient subgroups

Clinicians should control the <u>lens</u>, not the <u>model</u>

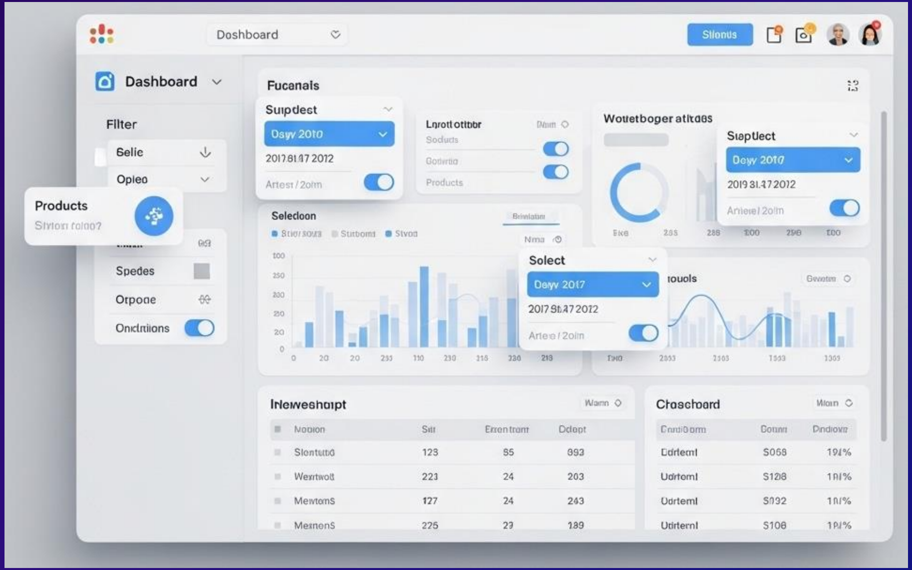
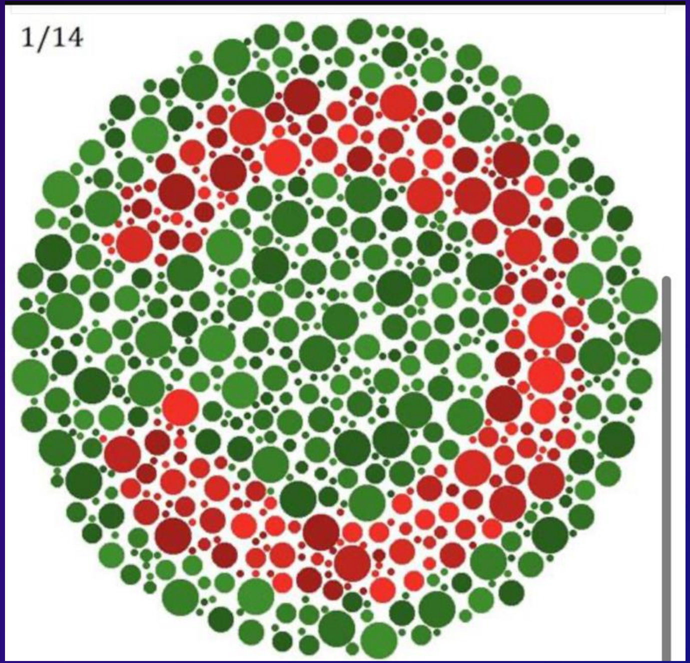


Image Generated by Gemini

Accessibility & Usability

- Font size, color blindness
- Mouse-free navigation
- Screen-reader compatibility



A11y Principles in Clinical Settings

Perceivable

Information and UI must be presented in ways users can perceive (regardless of abilities)

Operable

Users must be able to navigate and interact with interfaces regardless of physical ability

Understandable

Interfaces and content should be clear and predictable

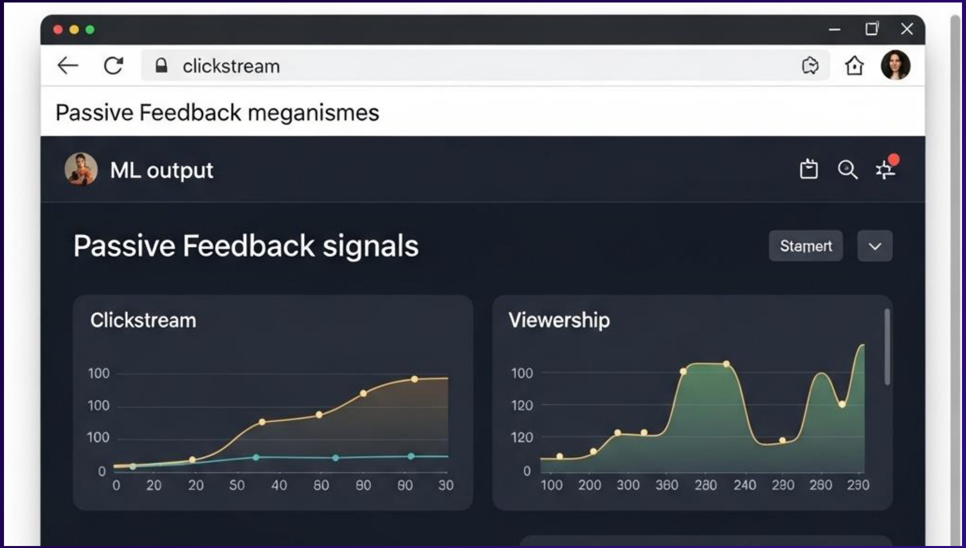
Robust

System must work with assistive technologies and remain functional across platforms and devices

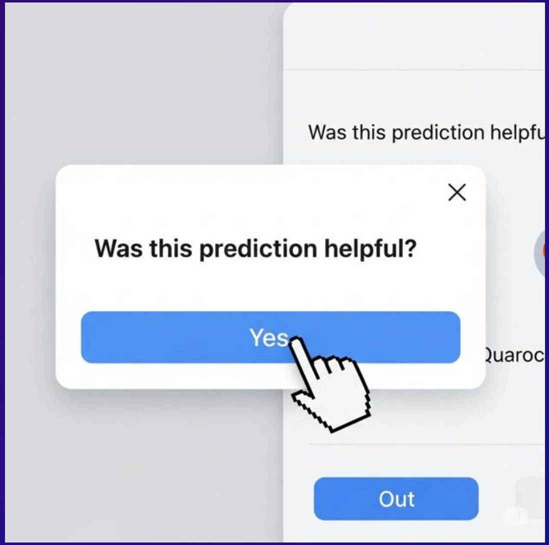
User Feedback Loops

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Passive Feedback



Active Feedback



What do you do with the clinical feedback?

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Testing Your UI with Clinicians

Ways to test it:

- Heuristic evaluation
- Think-aloud protocols
- Usability testing

Example metrics:

- System Usability Scale
- Task completion
- Cognitive load
- User experience

#6. User Training and Trust

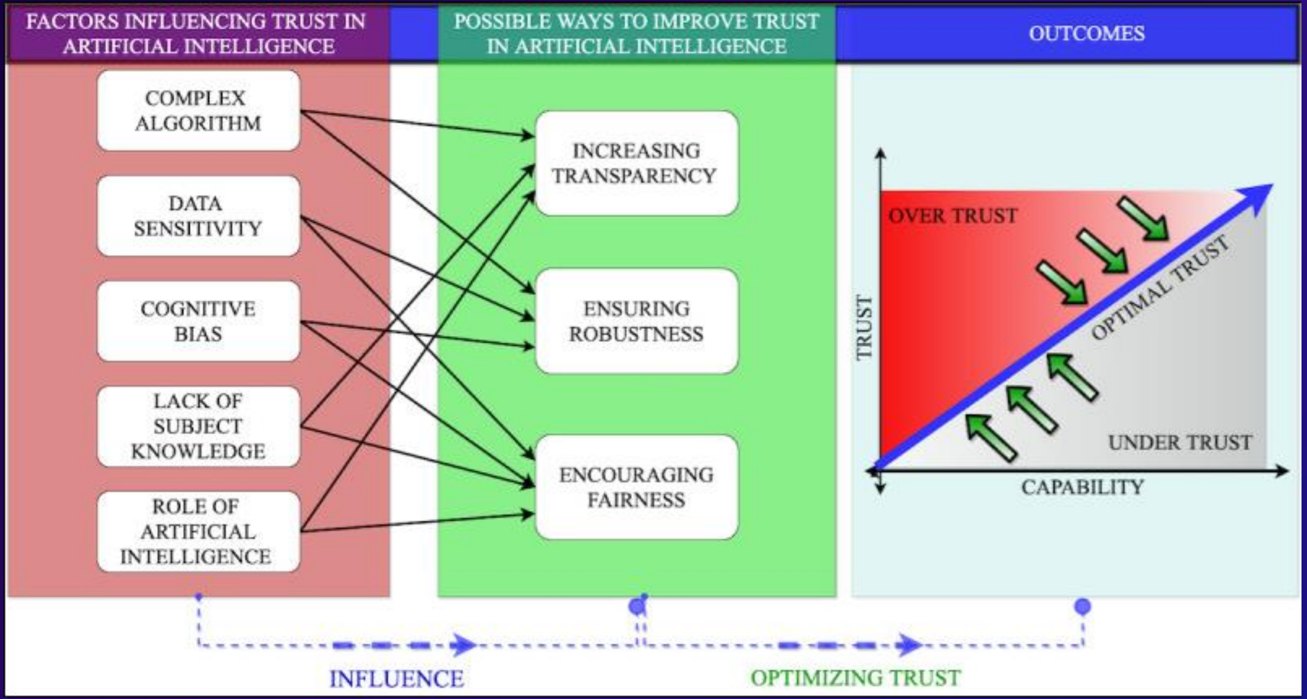
Model users require education, role modeling, and more

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Trust in AI

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Modified Material From: *AI Development and Implementation in Health Care **(LECUTRE:** End user interfacing with AI)*



Asan O, Bayrak AE, Choudhury A. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. J Med Internet Res. 2020 Jun 19;22(6):e15154. doi: 10.2196/15154. PMID: 32558657; PMCID: PMC7334754.

Trust in AI

What do Physician Value to Create Trust?

- Feature Importance
- Instance Level Explanations
- Uncertainty
- Temporal Explanations
- Transparent Design

Tonekaboni et al. 2019*. What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical end Use*

Applying this to our ROTEM case study

#6. User Training and Trust

Who do we mean by this?

What education do users need?
How can we foster trust and uptake?

#7. Reporting, regulation and liability considerations

Determining strategies for reporting feedback/events and liability

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Regulation

Not covering in detail due to country-specific differences

Canadian SaMD (Software as a Medical Device) Classification

State of healthcare situation or condition	Significance of information provided by SaMD to healthcare decision		
	Treat or diagnose	**Drive**	**Inform**
Critical	**III**	III	I or II
Serious	II or III	II or III	I or II
Non-serious	I or II	I or II	I or II

Regulation

Not covering in detail due to country-specific differences

World Health Organization

Health Topics | Countries | Newsroom | Emergencies

Home / News / WHO outlines considerations for regulation of artificial intelligence for health

WHO outlines considerations for regulation of artificial intelligence for health

19 October 2023 | News release | Geneva |Reading time: 2 min (610 words)

The World Health Organization (WHO) has released a new publication listing key regulatory considerations on artificial intelligence (AI) for health. The publication emphasizes the importance of establishing AI systems' safety and effectiveness, rapidly making appropriate

Model Governance

Key Questions

Ownership

Who owns (and is responsible for) the model after deployment?

Feedback

How do we capture end-user feedback?

Updates

How are they triggered and communicated

Best Practices

- Establish a model review board with clinical and technical experts
- Maintain changelogs and document rationale for changes
- Use version control

Model Facts Card

Sendak et al. 2020. npj digital medicine

Model Facts | **Model name:** Deep Sepsis | **Locale:** Duke University Hospital

Approval Date: 09/22/2019 | **Last Update:** 01/13/2020 | **Version:** 1.0

Summary

This model uses EHR input data collected from a patient's current inpatient encounter to estimate the probability that the patient will meet sepsis criteria within the next 4 hours. It was developed in 2016-2019 by the Duke Institute for Health Innovation. The model was licensed to Cohere Med in July 2019.

Mechanism

- **Outcome** .... sepsis within the next 4 hours, see outcome definition in "Other Information"
- **Output** .... 0% - 100% probability of sepsis occurring in the next 4 hours
- **Target population** .... all adult patients >18 y.o. presenting to DUH ED
- **Time of prediction** .... every hour of a patient's encounter
- **Input data source** .... electronic health record (EHR)
- **Input data type** .... demographics, analytes, vitals, medication administrations
- **Training data location and time-period** .... DUH, diagnostic cohort, 10/2014 – 12/2015
- **Model type** .... Recurrent Neural Network

Validation and performance

	Prevalence	AUC	PPV @ Sensitivity of 60%	Sensitivity @ PPV of 20%	Cohort Type	Cohort URL / DOI
Local Retrospective	18.9%	0.88	0.14	0.50	Diagnostic	arxiv.org/abs/1708.05894
Local Temporal	6.4%	0.94	0.20	0.66	Diagnostic	jmir.org/preprint/15182
Local Prospective	TBD	TBD	TBD	TBD	TBD	TBD
External	TBD	TBD	TBD	TBD	TBD	TBD
Target Population	6.4%	0.94	0.20	0.66	Diagnostic	jmir.org/preprint/15182

Uses and directions

- **Benefits:** Early identification and prompt treatment of sepsis can improve patient morbidity and mortality.
- **Target population and use case:** Every hour, data is pulled from the EHR to calculate risk of sepsis for every patient at the DUH ED. A rapid response team nurse reviews every high-risk patient with a physician in the ED to confirm whether or not to initiate treatment for sepsis.
- **General use:** This model is intended to be used to by clinicians to identify patients for further assessment for sepsis. The model is not a diagnostic for sepsis and is not meant to guide or drive clinical care. This model is intended to complement other pieces of patient information related to sepsis as well as a physical evaluation to determine the need for sepsis treatment.
- **Appropriate decision support:** The model identifies patient X as at a high risk of sepsis. A rapid response team nurse discusses the patient with the ED physician caring for the patient and they agree the patient does not require treatment for sepsis.
- **Before using this model:** Test the model retrospectively and prospectively on a diagnostic cohort that reflects the target population that the model will be used upon to confirm validity of the model within a local setting.
- **Safety and efficacy evaluation:** Analysis of data from clinical trial (NCT03655626) is underway. Preliminary data shows rapid response team, nurse-driven workflow was effective at improving sepsis treatment bundle compliance.

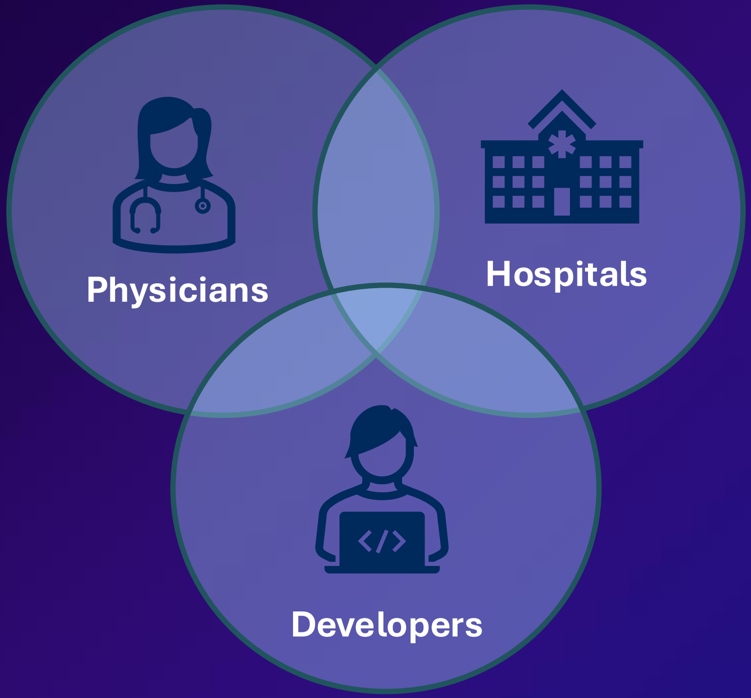
Warnings

- **Risks:** Even if used appropriately, clinicians using this model can misdiagnose sepsis. Delays in a sepsis diagnosis can lead to morbidity and mortality. Patients who are incorrectly treated for sepsis can be exposed to risks associated with unnecessary antibiotics and intravenous fluids.
- **Inappropriate Settings:** This model was not trained or evaluated on patients receiving care in the ICU. Do not use this model in the ICU setting without further evaluation. This model was trained to identify the first episode of sepsis during an inpatient encounter. Do not use this model after an initial sepsis episode without further evaluation.
- **Clinical Rationale:** The model is not interpretable and does not provide rationale for high risk scores. Clinical end users are expected to place model output in context with other clinical information to make final determination of diagnosis.
- **Inappropriate decision support:** This model may not be accurate outside of the target population, primarily adults in the non-ICU setting. This model is not a diagnostic and is not designed to guide clinical diagnosis and treatment for sepsis.
- **Generalizability:** This model was primarily evaluated within the local setting of Duke University Hospital. Do not use this model in an external setting without further evaluation.
- **Discontinue use if:** Clinical staff raise concerns about utility of the model for the indicated use case or large, systematic changes occur at the data level that necessitates re-training of the model.

Other information:

- **Outcome Definition:** https://doi.org/10.1101/648907
- **Related model:** http://doi.org/10.1001/jama.2016.0288
- **Model development & validation:** arxiv.org/abs/1708.05894
- **Model implementation**: jmir.org/preprint/15182
- **Clinical trial:** clinicaltrials.gov/ct2/show/NCT03655626
- **Clinical impact evaluation:** TBD
- **For inquiries and additional information:** please email mark.sendak@duke.edu

Liability Considerations



Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

Liability of Physicians

Liability of Physicians

Figure. Examples of Potential Legal Outcomes Related to AI Use in Clinical Practice

Scenario	AI recommendation	AI accuracy	Physician action	Patient outcome	Legal outcome (probable)
1	Standard of care	Correct	Follows	Good	No injury and no liability
2			Rejects	Bad	Injury and liability
3		Incorrect (standard of care is incorrect)	Follows	Bad	Injury but no liability
4			Rejects	Good	No injury and no liability
5	Nonstandard care	Correct (standard of care is incorrect)	Follows	Good	No injury and no liability
6			Rejects	Bad	*Injury but no liability*
7		Incorrect	Follows	Bad	*Injury and liability*
8			Rejects	Good	No injury and no liability

https://jamanetwork.com/journals/jama/article-abstract/2752750

Liability Considerations & UI

Facilitators of AI-Physician Handoff Elements:

- Transparency (who built it, when trained)
- Legal disclaimers and responsibility cues
- Documentation within the UI (include physician perspective)

Applying this to our case study

#7. Reporting, regulation and liability considerations

How should we provide a way for reporting?
Who should be responsible for reporting?
Who is liable for errors?

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

#8. Post-Deployment Monitoring

Ensuring our model continues to perform well while being used clinically

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

Why Post-Deployment is Critical

- Many models fail when deployed (despite strong development performance)
- Model behaviour can drift even if initially successful deployment
- Bias can emerge after deployment due to population or practice changes
- Clinicians rely on <u>ongoing</u> accuracy, trust, and relevance

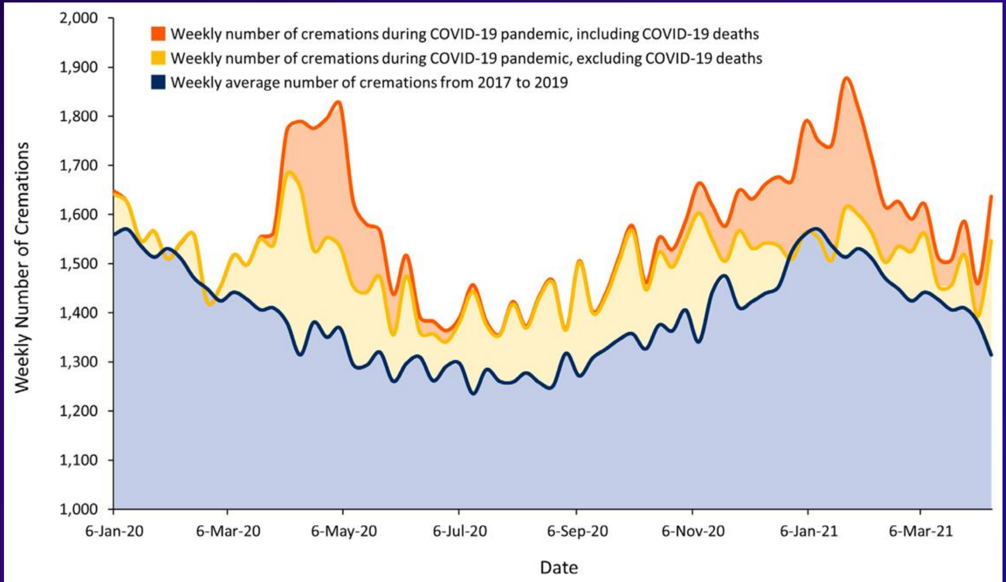
Performance Monitoring

Recommendations:

- Silent deployment / shadow deployment prior to full deployment
- Track key metrics continuously (e.g., AUROC, calibration, false positives)
- Compare across patient subgroups and time
- Consider using a live dashboard with trigger alerts if performance drops below a threshold

Data and Concept Drift

- **Data Drift**: change in input data distribution
- **Concept Drift**: change in outcome relationships



Postill et al. 2021 Science Brief

Model Drift and Data Shift

Detection Techniques

- **Statistical drift tests**: population stability index (PSI), KL divergence, KS test
- **Embedding drift** for NLP image data
- Monitor feature distribution and missingness over time
- Compare predictive vs. actual distributions regularly

Response Actions

- Retrain on updated data
- Use adaptive or online learning if applicable

What are We Doing in Toronto?

Image Generated by Gemini

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

Beyond Model Performance…

- Are clinical users interpreting the model correctly?
- Are model recommendations safe, usable, actionable?
- Any evidence of harm or unexpected consequences?

Beyond Model Performance...

Feedback Channels

In-tool rating prompts
Monthly clinician feedback meetings
Monitor override or ignore rates

Additional Metrics

Adoption rates
Action rates
Time-to-decision

Patient Engagement

Transparent reporting and integration of patient perspectives in post-deployment

Applying this to
our case study

#8. Post-Deployment Monitoring

How can we ensure that no alert doesn't mean the model is working?
Who is responsible?

04

Practical Exercise

Designing a UI for an ML model interpreting ROTEM

Temerty Centre for AI Research
and Education in Medicine
UNIVERSITY OF TORONTO

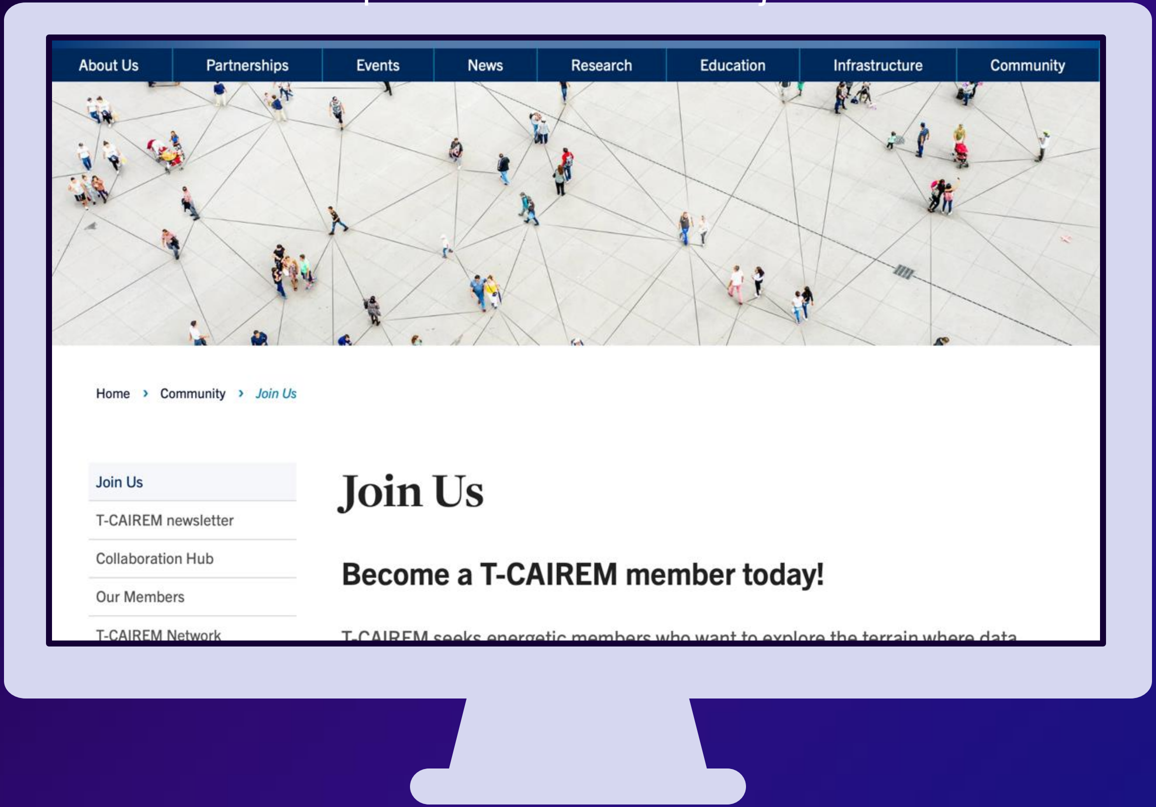
User Interface

Let’s imagine that we have addressed issues in model performance…

How could we design a UI to interpret ROTEM?

Join T-CAIREM

https://tcairem.utoronto.ca/join-us



Use our Health Data Nexus

Or contribute data to…

Workshop Committee





Thank You

Please contact me with any questions

gemma.postill@utoronto.ca

Gemma Postill
/gemma-p-80481987/

Temerty Centre for AI Research and Education in Medicine
UNIVERSITY OF TORONTO

CREDITS: This presentation template was created by **Slidesgo,** and includes icons by **Flaticon**, and infographics & images by **Freepik**