An Equity-Grounded Approach to Developing Computational Phenotypes for Peripheral Arterial Disease

Andrew Gonzalez, MD JD MPH
Assistant Professor of Surgery, Indiana School of Medicine
Associate Director for Data Science, Regenstrief Institute Center for Health Services Research

Disclosures

1. Research funding from Doris Duke Foundation, Indiana University Health Foundation, National Academy of Medicine
2. Stock and derivatives purchased on the open market (some bearish and some bullish positions) in
   – Apple, AMD, Palo Alto Networks, NVIDIA, Tesla, Google, Microsoft, Meta, Oracle,
3. Intellectual Bias
   – Patient-centered clinician autonomy

Delilah


Welcome to
Anushri's
Annaprashan

Goals of this talk

1. Motivation for DxEx project
2. What did I learn?
   - Technical / Tactical
   - Leadership / Strategy
3. Next Steps

Outline

Part 1 – (semi)-autonomous diagnosis in PAD

Part 2 – Experiential Learning

Part 3 – Next steps

Part 1 – **Diagnosis in PAD**

(Semi)-autonomous diagnosis in PAD

Part 1 – **Diagnosis in PAD**

Why can’t rely on human clinicians?

Humans alone aren't enough for PAD population health

1. Clinicians, **on average**, lack the knowledge, skill, inclination, and resources to reproducibly close the diagnostic loop **in PAD**
   - Knowledge
   - Skill (e.g. performing vascular ultrasound)
   - Inclination (e.g.)
   - Resources (e.g. vascular lab)
2. Patient awareness of PAD is low
   - Poor baseline knowledge
   - Poor retention
3. Patient-provider communication isn't *scalable*

EDUCATION CORNER

From the Canadian Society for Vascular Surgery

Knowledge gap of peripheral artery disease starts in medical school

Check for updates

Musaad AlHamzah, MD,[a,b] Rachel Eikelboom, MD,[c] Mohamad A. Hussain, MD,[a] Muzammil H. Syed, BSc(c),[d,e] Konrad Salata, MD,[a] Mark Wheatcroft, MBChB,[a,f] Subodh Verma, MD,[a,f,g] and Mohammed Al-Omran, MD,[a,d,g]
Toronto and London, Ontario, and Winnipeg, Manitoba, Canada; and Riyadh, Saudi Arabia

ABSTRACT

Objective: Previous data suggest that physicians have suboptimal knowledge about peripheral artery disease (PAD). Our aim was to evaluate Canadian medical students' knowledge of PAD to determine if this knowledge gap exists early in medical training.

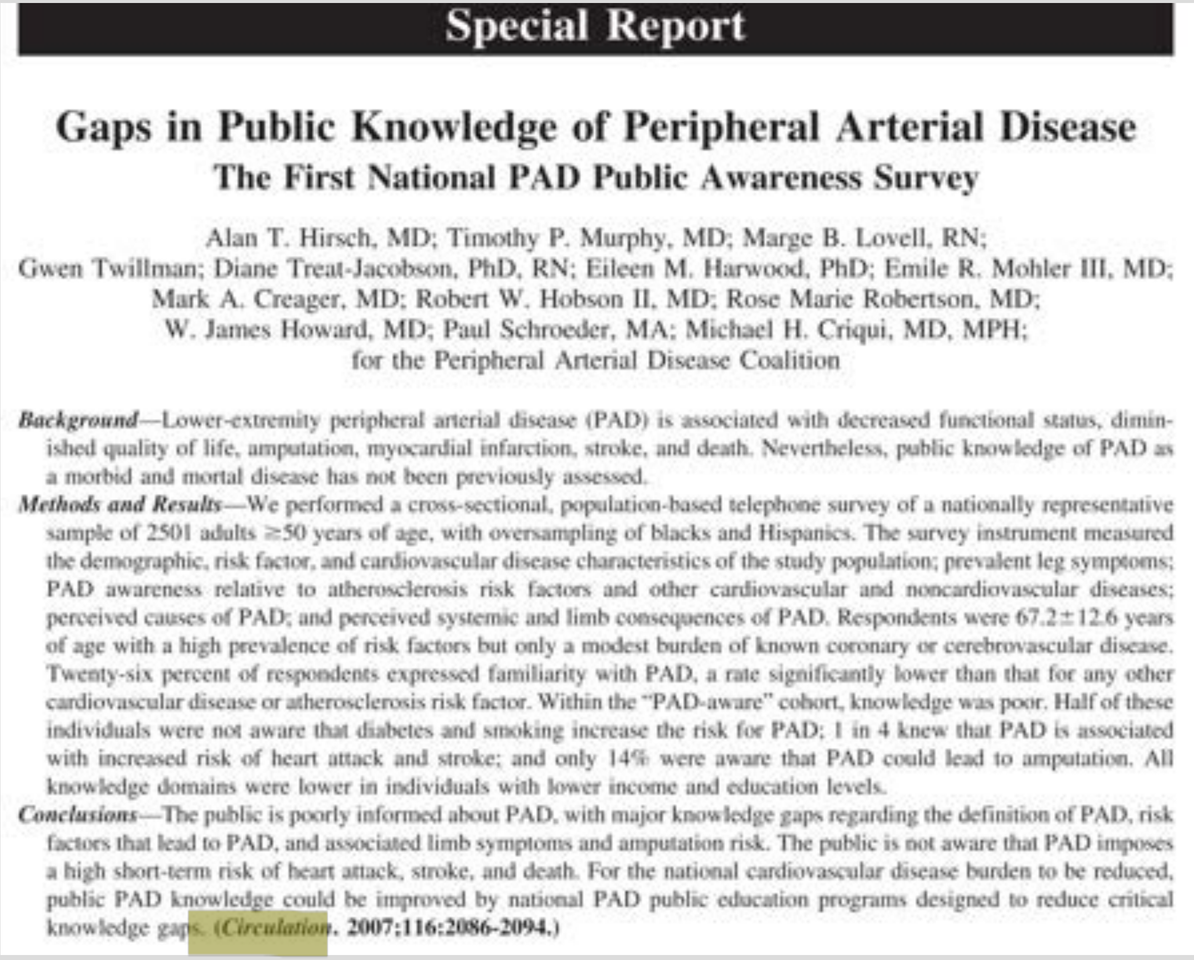
Methods: We conducted a descriptive, cross-sectional, interview-based study of graduating medical students at the University of Toronto. We used a standardized questionnaire to evaluate students' knowledge of PAD and coronary artery disease (CAD) in the following domains: clinical presentation, risk factors, preventative measures, treatment, and complications. We calculated mean (standard deviation [SD]) scores for each CAD and PAD knowledge domain and examined for differences in PAD vs CAD scores.

Results: Seventy-two graduating medical students participated in this study, of which females accounted for 58%. Nearly all participants reported being exposed to PAD (89%) and CAD (92%) through their medical school curriculum. Overall, medical students scored better in identifying CAD characteristics (mean [SD] score, 16.4 [2.7]) compared with PAD (mean [SD] score, 14.6 [3.2]) ($P < .0001$). This difference was driven by the inferior performance of students in identifying risk factors ($P < .0001$), preventative measures ($P = .049$), and complications ($P < .0001$) of PAD compared with CAD. Out-of-class exposure (eg, clinical rotation, research experience) had a positive impact on students knowledge of both PAD and CAD.

Conclusions: Our results demonstrate suboptimal knowledge of medical graduates of both CAD and PAD. Although they share common atherosclerotic risk factors and cardiovascular complications, medical students were less likely to associate these with PAD than CAD. We recommend a comprehensive module that incorporates all presentations of atherosclerotic disorders to enhance students' understanding of these pathologies in medical schools. (J Vasc Surg 2019;70:241-5.)

Humans alone aren't enough for PAD population health

1. Clinicians, **on average**, lack the knowledge, skill, inclination, and resources to reproducibly close the diagnostic loop <u>**in PAD**</u>
   - Knowledge
   - Skill (e.g. performing vascular ultrasound)
   - Inclination (e.g.)
   - Resources (e.g. vascular lab)
2. Patient awareness of PAD is low
   - Poor baseline knowledge
   - Poor retention
3. Patient-provider communication isn't *scalable*



Special Report

Gaps in Public Knowledge of Peripheral Arterial Disease
The First National PAD Public Awareness Survey

Alan T. Hirsch, MD; Timothy P. Murphy, MD; Marge B. Lovell, RN; Gwen Twillman; Diane Treat-Jacobson, PhD, RN; Eileen M. Harwood, PhD; Emile R. Mohler III, MD; Mark A. Creager, MD; Robert W. Hobson II, MD; Rose Marie Robertson, MD; W. James Howard, MD; Paul Schroeder, MA; Michael H. Criqui, MD, MPH; for the Peripheral Arterial Disease Coalition

Background—Lower-extremity peripheral arterial disease (PAD) is associated with decreased functional status, diminished quality of life, amputation, myocardial infarction, stroke, and death. Nevertheless, public knowledge of PAD as a morbid and mortal disease has not been previously assessed.
Methods and Results—We performed a cross-sectional, population-based telephone survey of a nationally representative sample of 2501 adults ≥50 years of age, with oversampling of blacks and Hispanics. The survey instrument measured the demographic, risk factor, and cardiovascular disease characteristics of the study population; prevalent leg symptoms; PAD awareness relative to atherosclerosis risk factors and other cardiovascular and noncardiovascular diseases; perceived causes of PAD; and perceived systemic and limb consequences of PAD. Respondents were 67.2±12.6 years of age with a high prevalence of risk factors but only a modest burden of known coronary or cerebrovascular disease. Twenty-six percent of respondents expressed familiarity with PAD, a rate significantly lower than that for any other cardiovascular disease or atherosclerosis risk factor. Within the "PAD-aware" cohort, knowledge was poor. Half of these individuals were not aware that diabetes and smoking increase the risk for PAD; 1 in 4 knew that PAD is associated with increased risk of heart attack and stroke; and only 14% were aware that PAD could lead to amputation. All knowledge domains were lower in individuals with lower income and education levels.
Conclusions—The public is poorly informed about PAD, with major knowledge gaps regarding the definition of PAD, risk factors that lead to PAD, and associated limb symptoms and amputation risk. The public is not aware that PAD imposes a high short-term risk of heart attack, stroke, and death. For the national cardiovascular disease burden to be reduced, public PAD knowledge could be improved by national PAD public education programs designed to reduce critical knowledge gaps. (*Circulation*. 2007;116:2086-2094.)

Humans alone aren’t enough for PAD population health

1. Clinicians, **on average**, lack the knowledge, skill, inclination, and resources to reproducibly close the diagnostic loop **in PAD**
   - Knowledge
   - Skill (e.g. performing vascular ultrasound)
   - Inclination (e.g.)
   - Resources (e.g. vascular lab)
2. Patient awareness of PAD is low
   - Poor baseline knowledge
   - Poor retention
3. Patient-provider communication isn’t *scalable*

VASCULAR MEDICINE

Review Article

Vascular Medicine
2020, Vol. 25(3) 263–273

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1358863X19893003
journals.sagepub.com/home/vmj
SAGE

Knowledge of peripheral artery disease: What do the public, healthcare practitioners, and trainees know?

Bernadeta M Bridgwood[1], Andrew TO Nickinson[1], John SM Houghton[1], Coral J Pepper[2] and Rob D Sayers[1]

Abstract

This systematic review evaluated the knowledge and awareness of peripheral artery disease (PAD) within the general public (including patients with peripheral vascular disease), nonspecialist healthcare professionals (nsHCP), and trainees (medical students and trainee doctors). Relevant articles were identified from electronic databases using key search terms: 'peripheral artery disease'; 'limb ischaemia'; 'intermittent claudication'; 'knowledge'; 'understanding'; 'public'; 'medical professional'. The heterogeneous results were described narratively. A lack of knowledge and understanding of PAD (disease awareness) were identified in all groups. Among nsHCPs, factors which affect knowledge include the level of training, early clinical exposure and the presence of family members with cardiovascular/vascular disease. Within the general public, knowledge and awareness was improved if a family member/friend had a diagnosis, or following a patient-centred consultation with any HCP. Public campaigns are proven effective in improving disease knowledge/awareness in conditions such as stroke alongside sustained patient education. These may provide future avenues to improve PAD knowledge and awareness, in order to effectively manage risk factors and minimise delayed or missed diagnosis of PAD. (**PROSPERO registration number: CRD42018117304**)

Humans alone aren’t enough for PAD population health

1. Clinicians, **on average**, lack the knowledge, skill, inclination, and resources to reproducibly close the diagnostic loop **<u>in PAD</u>**
   - Knowledge
   - Skill (e.g. performing vascular ultrasound)
   - Inclination (e.g.)
   - Resources (e.g. vascular lab)
2. Patient awareness of PAD is low
   - Poor baseline knowledge
   - Poor retention
3. Patient-provider communication isn’t *scalable*

High ●
Moderate ◉
Low ○

Knowledge level	
◉	Suboptimal PAD knowledge in medical graduates
◉	Lack of clarity of referral guidelines
◉	No trainee knew how to completely perform an ABI
○	Knowledge of PAD was 60.5% and less known about symptoms/medication
◉	Improved PAD knowledge with education
●	Uncertainty and varied PAD beliefs
○	Poor PAD knowledge and ABI technique
●	Lack of awareness of PAD
●	General participant PAD knowledge
◉	Lack of training, time, equipment and skills for PAD
●	Mixed knowledge and awareness
◉	Poor knowledge with varied reported training
○	Knowledge of risk factors varied between 44.7% and 71.8%
○	No knowledge increase with knowledge only-based education

https://pubmed.ncbi.nlm.nih.gov/32000617/

Humans alone aren’t enough for PAD population health

1. Clinicians, **on average**, lack the knowledge, skill, inclination, and resources to reproducibly close the diagnostic loop **<u>in PAD</u>**
   - Knowledge
   - Skill (e.g. performing vascular ultrasound)
   - Inclination (e.g.)
   - Resources (e.g. vascular lab)
2. Patient awareness of PAD is low
   - Poor baseline knowledge
   - Poor retention
3. Patient-provider communication isn’t *scalable*

High ● Moderate ◉ Low ○	Knowledge level
○	PAD knowledge depends on health insurance status
◉	Largely unaware of PAD
●	Gender differences in claudication description
○	Poor PAD and risk factor knowledge
◉	Low levels of PAD knowledge and awareness
◉	Risk factor awareness for PAD is suboptimal
●	PAD knowledge was poor
○	Specialist consultation improves PAD knowledge
○	80% correctly responded regarding foot care
○	Low awareness of PAD symptoms/risk factors
●	83% of patients with prior PAD knew their diagnosis
●	nsHCP PAD awareness is low
◉	All groups performed poorly

https://pubmed.ncbi.nlm.nih.gov/32000617/

Clinical interviews are inherently non-*scalable* due to structural aspects of communication

1. Patient-provider communication is often multi-turn

Clinical interviews are <u>inherently non-*scalable*</u> due to structural aspects of communication

1. Patient-provider communication is often <u>multi-turn</u>
2. Very **limited opportunities for $\mathbf{1{:}}\boldsymbol{m}_{\textbf{people}}$** communication are <u>inversely proportional to the conversation's depth/sensitivity</u>.

Clinical interviews are inherently non-*scalable* due to structural aspects of communication

1. Patient-provider communication is often multi-turn
2. Very **limited opportunities for** $\mathbf{1{:}}\boldsymbol{m}_{\mathbf{people}}$ communication are inversely proportional to the conversation's depth/sensitivity.

Clinical Scenario	1:m Comm Opp	Depth of Comm
Preoperative bariatric surgery information session conducted by a nurse	High	Shallow
taking a typical history and physical for a geriatric patient with her daughter	Moderate	Medium
Taking a focused history from a teen who you suspect is being abused by family member	Very low	Deep

Clinical interviews are <u>inherently non-*scalable*</u> due to structural aspects of communication

Clinical Scenario	1:m Comm Opp	Depth of Comm
Preoperative bariatric surgery information session conducted by a nurse	High	Shallow
taking a typical history and physical for a geriatric patient with her daughter	Moderate	Medium
Taking a focused history from a teen who you suspect is being abused by family member	Very low	Deep

3. Filtered through various kinds of "**non-random error** associated with increased delta between reality and perception"

Clinical interviews are inherently non-*scalable* due to structural aspects of communication

1. Patient-provider communication is often multi-turn
2. Very **limited opportunities for $1{:}m_{\text{people}}$** communication are inversely proportional to the conversation's depth/sensitivity.

Clinical Scenario	1:m Comm Opp	Depth of Comm
Preoperative bariatric surgery information session conducted by a nurse	High	Shallow
taking a typical history and physical for a geriatric patient with her daughter	Moderate	Medium
Taking a focused history from a teen who you suspect is being abused by family member	Very low	Deep

3. Filtered through various kinds of "**non-random error** associated with increased delta between reality and perception"

We need **human-machine** hybrid systems…

Part 2 – **Experiential Learning**

Focusing on systems

Hierarchical components of a clinical *socio-technical* system

- {**Service**} (class: Clinical, research, administrative, legal, logistical, business)
  - Clinical {**activity**} composed of **Tasks** (thinking, doing, recording, communicating) with **Purposes** (*diagnosis*, *prognosis*, *treatment selection*)
    - Data processing {**pipeline**}
      - machine {**agents**}
        - Deep learning {**models**}

Hierarchical components of a clinical *socio-technical* system

1. {**Service**} (class: Clinical, research, administrative, legal, logistical, business)
2. Clinical {**activity**}
   - **Task** (output): thinking, doing, recording, communicating
   - **Purpose** (1st principles): *diagnosis*, *prognosis*, *treatment selection*
3. Data processing {**pipeline**}
4. machine {**agents**}
5. Deep learning {**models**}

Hierarchical components of a <u>clinical</u> *socio-technical* system

1. {**Service**} (class: Clinical, research, administrative, legal, logistical, business)
2. Clinical {**activity**}
   - **Task** (output): <u>thinking</u>, <u>doing</u>, <u>recording</u>, <u>communicating</u>
   - **Purpose** (1st principles): *diagnosis*, *prognosis*, *treatment selection*
3. Data processing {**pipeline**}
4. machine {**agents**}
5. Deep learning {**models**}

Hierarchical components of a clinical *socio-technical* system

1. {**Service**} (class: Clinical, research, administrative, legal, logistical, business)
2. Clinical {**activity**}
   - **Task** (output): thinking, doing, recording, communicating
   - **Purpose** (1st principles): *diagnosis*, *prognosis*, *treatment selection*
3. Data processing {**pipeline**}
4. machine {**agents**}
5. Deep learning {**models**}

Part 2 – **Experiential Learning**

iCHAI DECODE framework

for *developing & deploying*

clinical AI systems

iCHAI Conceptual model {***end-to-end*** **ML Dev-Dep pipeline**}

D	E	C	O	D	E
Define	**Extract**	**Compute**	**Optimize**	**Deploy**	**Evaluate**

iCHAI Conceptual model **{*end-to-end* ML Dev-Dep pipeline}**

Data Level		*Model Level*		*Implementation Level*	
D	E	C	O	D	E
Define	**Extract**	**Compute**	**Optimize**	**Deploy**	**Evaluate**

iCHAI Conceptual model {*end-to-end* **ML Dev-Dep pipeline**}

Data Level | *Model Level* | *Implementation Level*

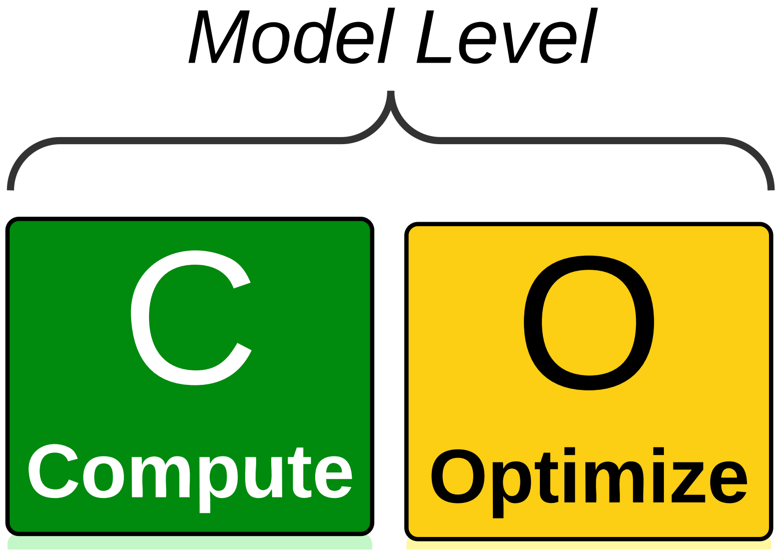
D	E	C	O	D	E
Define	**Extract**	**Compute**	**Optimize**	**Deploy**	**Evaluate**
Clinicians, patients and caretakers define a ***real-world problem and existing clinical constraints*** *to provide boundary conditions for potential solutions.*	*Clinicians and machine learning scientists collaborate to* ***identify and clean required data substrates*** *for model ingestion*	*Clinicians and machine learning scientists develop a* ***task-focused minimal viable model(s)*** *for technical and clinical validation.*	*Models are optimized, safety-checked, and packaged for pilot deployment as a* ***minimally viable product****.*	*Models are deployed as a* ***minimally viable experiment*** *to evaluate real world performance by end-users who were not part of the development team.*	*Models undergo* ***rigorous and systematic assessment*** *of quality (accuracy, alignment), usability, equity, and sustainability.*

Feedback

iCHAI Conceptual model {***end-to-end*** **ML Dev-Dep pipeline**}

Data Level		*Model Level*		*Implementation Level*	
D	E	C	O	D	E
Define	**Extract**	**Compute**	**Optimize**	**Deploy**	**Evaluate**

{*end-to-end* ML Dev-Dep pipeline}
is modular & model agnostic



{*end-to-end* ML Dev-Dep pipeline}
is modular & model agnostic

🔥 swapable

Model Level

C
Compute

O
Optimize

{*end-to-end* ML Dev-Dep pipeline}
is modular & model agnostic

🔥 swapable

OpenAI

Model Level

C
Compute

O
Optimize

{*end-to-end* ML Dev-Dep pipeline}
is modular & model agnostic

swapable

Model Level

C
Compute

O
Optimize

{*end-to-end* ML Dev-Dep pipeline}
is modular & model agnostic

swapable

Google Gemini

Model Level

C
Compute

O
Optimize

{*end-to-end* ML Dev-Dep pipeline}
is modular & model agnostic

🔥 swapable

deepseek

Model Level

C
Compute

O
Optimize

{*end-to-end* ML Dev-Dep pipeline}
is modular & model agnostic

🔥 swapable

Claude BY ANTHROP\C

Model Level

C
Compute

O
Optimize

INDIANA UNIVERSITY SCHOOL OF MEDICINE
Department of Surgery | Division of Vascular Surgery

REGENSTRIEF INSTITUTE
Center for Health Services Research

{*end-to-end* ML Dev-Dep pipeline}
is modular & model agnostic

🔥 swapable

Google Gemini

Model Level

C
Compute

O
Optimize

{*end-to-end* ML Dev-Dep pipeline}
is modular & model agnostic

swapable

MISTRAL AI_

Model Level

C
Compute

O
Optimize

{*end-to-end* ML Dev-Dep pipeline}
is modular & model agnostic

swapable

Model Level

C
Compute

O
Optimize

Part 2 – **Experiential Learning**

A framework for real-world performance

Which domains of performance matter in the **real world**?

Original Evaluation Framework

Table 1. iCHAI E^3 Evaluation framework for clinical AI systems

	Data	*Deep Learning Model* (or chained system)	*Downstream Task* (Clinical or pop. Health)
Efficacy			
Efficiency			
Equity			

Original Evaluation Framework

Table 1. iCHAI E^3 Evaluation framework for clinical AI systems

		Data	**Deep Learning Model** (or Chained System)	**Downstream Task** (clinical or pop. health)
Efficacy	*Discussion*			
	Experiments			
	Deliverable			
Efficiency	*Discussion*			
	Experiments			
	Deliverable			
Equity	*Discussion*			
	Experiments			
	Deliverable			

Original Evaluation Framework

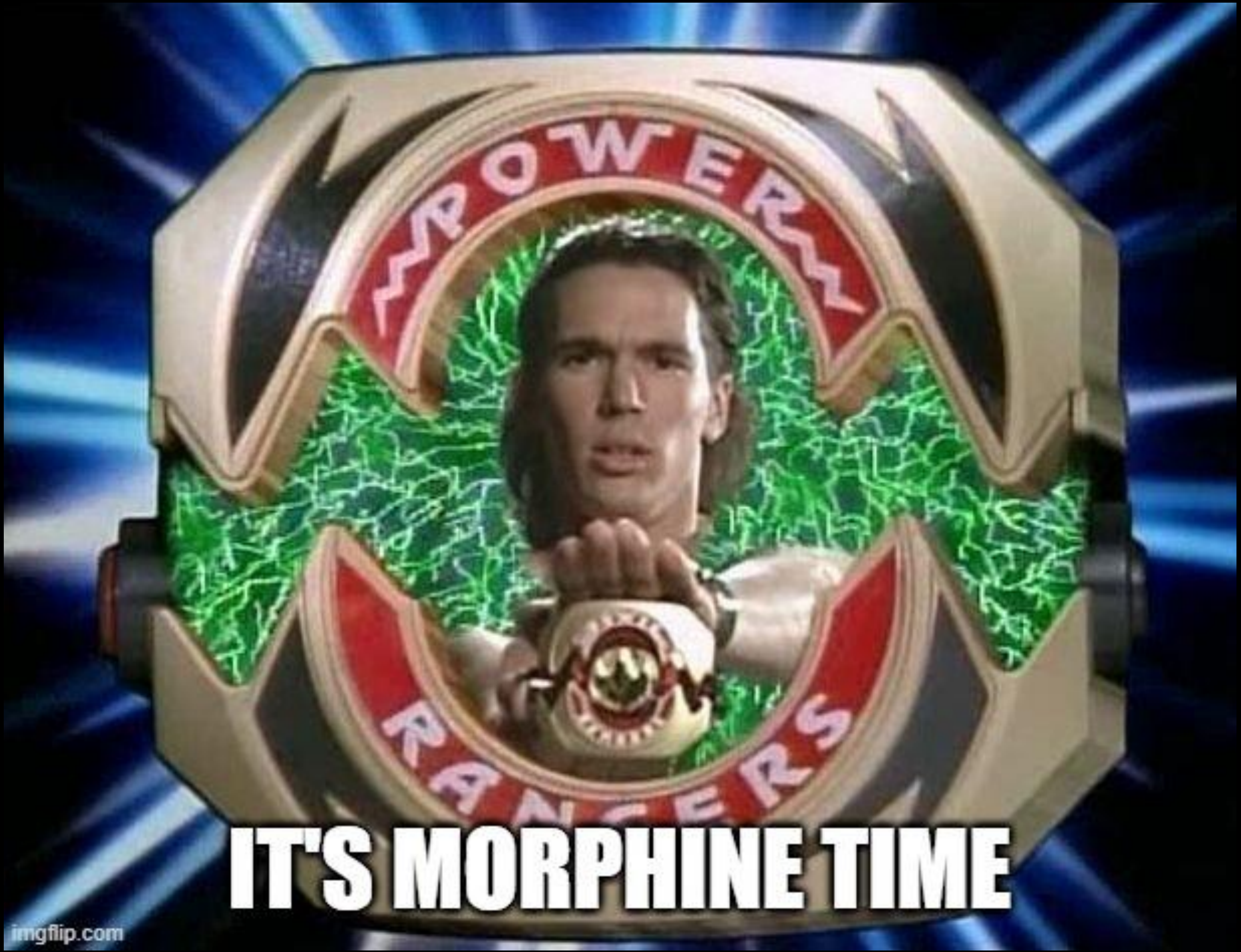
Table 1. iCHAI E^3 Evaluation framework for clinical AI systems

	Data	*Deep Learning Model* (or chained system)	*Downstream Task* (Clinical or pop. Health)
Efficacy			
Efficiency			
Equity			

Original Evaluation Framework

Table 1. iCHAI E^3 Evaluation framework for clinical AI systems

	Data	***Deep Learning Model*** *(or chained system)*	***Downstream Task*** *(Clinical or pop. Health)*
Efficacy			
Efficiency			
Equity			


POWER
RANGERS
IT'S MORPHINE TIME
imgflip.com

Revised Evaluation Framework

Table 1. iCHAI A2 Evaluation framework for clinical AI system minimal viability

	Input Data	*Output Predictions*	
		Deterministic	*Generative*
Accuracy	**Observability** of the training data? **Transparency** labelling ontology? **Provenance** of the labeling process?	Can we **verify** if given the data the model is attending to? (e.g. a text snippet)	Would this **change clinical management**? (decision focused)
Alignment	In the real-world, is this how the input data will look with respect to format and content?	Format and content you want to see at the point of care? Do providers make better decisions in response to this digital health intervention?	

Improving performance – let's take a trip to the dojang

{*end-to-end* ML Dev-Dep pipeline}
is modular & model agnostic

swapable

{Model}

Model Level

C
Compute

O
Optimize

{*end-to-end* ML Dev-Dep pipeline} is amenable to **clinical-technical optimization**

🔥 swapable

{Model}

Model Level

C
Compute

O
Optimize

🤦‍♀️ in the loop

{*end-to-end* ML Dev-Dep pipeline}
is amenable to **clinical-technical optimization**

🔥 swapable

{Model}

Model Level

C
Compute

O
Optimize

🤦‍♀️ in the loop

1. More Data
2. Prompt Tuning
3. Hyperparameter Optimization
4. Fine Tuning

{*end-to-end* ML Dev-Dep pipeline} is amenable to **clinical-technical optimization**

O
Optimize

🤦🏽‍♀️ in the

Figure. Iterative "step-up approach" to improving Vision Language Model (VLM) performance in a Retrieval Augmented Generation (RAG) based system

	Capable of Implementing?			
	End-user alone	*Coder alone*	*Clinical SME alone*	***Coder + SME***
Step 7 **Parameter efficient fine-tuning**	No	Maybe	No	Yes
Step 6 **Graph-RAG Application**	No	Maybe	No	Yes
Step 5 **Active Learning strategies**	No	Maybe	No	Yes
Step 4 **Instruction tuning** (extractive) **or re-ranking** (summaries)	Maybe	No	Yes	Yes
Step 3 **"Few shot" in-context learning**	Maybe	No	Yes	Yes
Step 2 **Link LLM to knowledge base on disease (e.g. Pubmed)**	No	Maybe	Maybe	Yes
Step 1 **Hyperparameter modification**	Yes	Maybe	Yes	Yes

O
Optimize

{*end-to-end* ML Dev-Dep pipeline}
is amenable to **clinical-technical optimization**

🤦🏽‍♀️ in the

Figure. Iterative "step-up approach" to improving Vision Language Model (VLM) performance in a Retrieval Augmented Generation (RAG) based system

	Capable of Implementing?			
	End-user alone	*Coder alone*	*Clinical SME alone*	*Coder + SME*
Step 7 **Parameter efficient fine-tuning**	No	Maybe	No	Yes
Step 6 **Graph-RAG Application**	No	Maybe	No	Yes
Step 5 **Active Learning strategies**	No	Maybe	No	Yes
Step 4 **Instruction tuning** (extractive) **or re-ranking** (summaries)	Maybe	No	Yes	Yes
Step 3 **"Few shot" in-context learning**	Maybe	No	Yes	Yes
Step 2 **Link LLM to knowledge base on disease (e.g. Pubmed)**	No	Maybe	Maybe	Yes
Step 1 **Hyperparameter modification**	Yes	Maybe	Yes	Yes

{*end-to-end* ML Dev-Dep pipeline} is amenable to **clinical-technical optimization**

O
Optimize

in the

Figure. Iterative "step-up approach" to improving Vision Language Model (VLM) performance in a Retrieval Augmented Generation (RAG) based system

	Capable of Implementing?			
	End-user alone	*Coder alone*	*Clinical SME alone*	***Coder + SME***
Step 7 **Parameter efficient fine-tuning**	No	Maybe	No	Yes
Step 6 **Graph-RAG Application**	No	Maybe	No	Yes
Step 5 **Active Learning strategies**	No	Maybe	No	Yes
Step 4 **Instruction tuning** (extractive) **or re-ranking** (summaries)	Maybe	No	Yes	Yes
Step 3 **"Few shot" in-context learning**	Maybe	No	Yes	Yes
Step 2 **Link LLM to knowledge base on disease (e.g. Pubmed)**	No	Maybe	Maybe	Yes
Step 1 **Hyperparameter modification**	Yes	Maybe	Yes	Yes

{*end-to-end* ML Dev-Dep pipeline}
is amenable to **clinical-technical optimization**

O
Optimize

🤦🏽‍♀️ in the loop

Figure. Iterative "step-up approach" to improving Vision Language Model (VLM) performance in a Retrieval Augmented Generation (RAG) based system	Capable of Implementing? *End-user alone*	*Coder alone*	*Clinical SME alone*	***Coder + SME***
Step 7 **Parameter efficient fine-tuning**	No	Maybe	No	Yes
Step 6 **Graph-RAG Application**	No	Maybe	No	Yes
Step 5 **Active Learning strategies**	No	Maybe	No	Yes
Step 4 **Instruction tuning** (extractive) **or re-ranking** (summaries)	Maybe	No	Yes	Yes
Step 3 **"Few shot" in-context learning**	Maybe	No	Yes	Yes
Step 2 **Link LLM to knowledge base on disease (e.g. Pubmed)**	No	Maybe	Maybe	Yes
Step 1 **Hyperparameter modification**	Yes	Maybe	Yes	Yes

Part 2 – **Experiential Learning**

Current Status

Association for Academic Surgery

Use of Deep Learning to Identify Peripheral Arterial Disease Cases From Narrative Clinical Notes

Shantanu Dev, BS,[a,b] Andrew Zolensky, BS,[b,c] Hanaa Dakour Aridi, MD,[d] Catherine Kelty, PhD, MS,[e] Mackenzie K. Madison, MD, MS,[d] Anush Motaganahalli, MPH,[b] Benjamin S. Brooke, MD, PhD, FACS,[f] Brian Dixon, PhD, MPA,[g] Malaz Boustani, MD, MPH,[h,i] Zina Ben Miled, PhD,[j] Ping Zhang, PhD,[a] and Andrew A. Gonzalez, MD, JD, MPH, FACS[b,d,k,]*

[a] Department of Computer Science and Engineering, College of Engineering, The Ohio State University, Columbus, Ohio
[b] Center for Health Services Research, Regenstrief Institute, Indianapolis, Indiana
[c] Department of Computer and Information Science, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, Pennsylvania
[d] Division of Vascular Surgery, Indiana University School of Medicine, Indianapolis, Indiana
[e] Division of Nephrology, Indiana University School of Medicine, Indianapolis, Indiana
[f] Department of Surgery, Utah Intervention Quality & Implementation Research Group (U-INQUIRE), University of Utah, Salt Lake City, Utah
[g] Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, Indiana
[h] Center for Health Innovation & Implementation Science, Indiana University, Indianapolis, Indiana
[i] Center for Aging Research, Regenstrief Institute, Indianapolis, Indiana
[j] Department of Electrical and Computer Engineering, Lemar Institute of Technology, Beaumont, Texas
[k] Department of Surgery, Surgical Outcomes & Quality Improvement Center (SOQIC), Indiana University School of Medicine, Indianapolis, Indiana

Table 1 – Study demographics, stratified by training/testing cohort for the Indiana Peripheral Arterial Disease Cohort (2010-2020).

Variables	Training set				Testing set			
	Total	%	PAD codes	Non-PAD codes	Total	%	PAD codes	Non-PAD codes
Race								
White	229,731	67.8%	13,114	216,617	98,892	68.1%	5505	93,387
Black	99,096	29.2%	4513	94,583	41,981	28.9%	1924	40,057
Other	10,222	3.0%	296	9926	4441	3.0%	132	4309
Ethnicity								
Hispanic or Latino %	18,238	5.4%	608	17,630	7845	5.6%	260	7585
Not Hispanic or Latino %	312,343	92.1%	16,308	296,035	133,912	92.2%	6895	127,017
Unknown ethnicity %	8468	2.5%	1011	7457	3557	2.4%	406	3151
Gender								
Male %	168,960	49.8%	10,208	158,752	72,759	50.1%	4300	68,459
Female %	170,089	50.2%	7715	162,374	72,555	49.9%	3261	69,294

Table 2 – Model performance on the testing dataset for PAD note classification task from the Indiana Peripheral Arterial Disease Cohort (2010-2020), stratified by search strategy.

Performance metric	Keyword search	DL	P value
Mathew correlation coefficient	0.20	0.51	<0.001
Sensitivity	0.62	0.70	<0.001
Specificity	0.94	0.99	<0.001
Positive predictive value	0.69	0.82	<0.001
Negative predictive value	0.96	0.97	<0.001
Accuracy	0.91	0.96	<0.001

At the encounter level, ground truth labels for PAD encounters *versus* non-PAD encounters is based on ICD 9/10 codes. Meaning, positive cases are those with PAD ICD-9/10 codes and negative controls are those without PAD ICD-9/10 codes for a given encounter.

Table 4 – Examples of false positive and false negative model predictions using the BioMed-RoBERTa language model on the Indiana Peripheral Arterial Disease Cohort (2010-2020).

Incorrect response type	Ref	Note type	Text for discussion	Comment regarding potential misclassification reasons
False negative examples	1	Discharge summary	With the goal of trying to avoid the surgery.... The rest of the hospital stay the patient was followed as appropriate by ...cardiovascular surgery. As it turned out... did not need any cardiovascular surgery interventions.	Model confused by the similar verbiage between vascular and cardiac surgery.
	2	Discharge summary	Past medical history: 1. coronary artery disease. 2. peripheral vascular disease... [CT abdomen and pelvis w/IV contrast] ..."vasculature: Advanced atherosclerotic change of the abdominal aorta without aneurysm"	Artifact from billing code for PAD being present despite PAD treatment not being included in the prose of this discharge summary.

Studies to evaluate “**non-random error** associated with increased delta between reality and perception”

1. **Ablation**
   - Systematically remove demographics and evaluate model performance
   - For generative tasks vary demographics and see if prediction changes

Studies to evaluate "**non-random error** associated with increased delta between reality and perception"

1. **Ablation**
   - Systematically remove demographics and evaluate model performance
   - For generative tasks vary demographics and see if prediction changes
2. **Reverse prediction**
   - Predict demographics instead of outcomes

Studies to evaluate “**non-random error** associated with increased delta between reality and perception”

1. **Ablation**
   – Systematically remove demographics and evaluate model performance
   – For generative tasks vary demographics and see if prediction changes
2. **Reverse prediction**
   – Predict demographics instead of outcomes
3. **Differential performance**
   – Our model, your data
   – Our data, your model

Working Titles	H_0 testing	Reviews	Viewpoints
Impact of race, gender, and age on LLM based diagnosis of Peripheral Arterial Disease from provider narratives	X		
Assisting sources of error in machine-based diagnosis of Peripheral Arterial Disease	X		
***Adaptive Performance Benchmarking in Healthcare Foundational Models**: the need for a national assurance lab*			X
***Beyond Explainability** – How to build trustworthy systems in clinical machine learning*		X	
The culture divide between medicine and Silicon Valley: *causes, implications, and a path forward*			X
Escaping the AI Hype Cycle: refocusing healthcare AI on problems that matter			X
A {roadmap, playbook} for successful academic industry partnerships in healthcare AI		X	

Part 3 – **Next steps**

Funding to impact

Subproblems with different boundary conditions:

}

$$\lim_{0 \to 1} v(x) = \boldsymbol{viability}$$

}

Subproblems with <u>different boundary conditions</u>:

- $$\lim_{0\to 1} v(x) = \boldsymbol{viability}$$
- $$\lim_{1\to 100} s(y) = \boldsymbol{scale}$$
- $$v(x) \neq s(y)$$

Subproblems with <u>different boundary conditions</u>:

} $\lim\limits_{0\to1} v(x) = \boldsymbol{viability}$

} $\lim\limits_{1\to100} s(y) = \boldsymbol{scale}$

} $v(x) \neq s(y)$

} **pilot** $v(x)$,

```
if viable >1, get($)
else, die trying
```

Subproblems with <u>different boundary conditions</u>:

- $\lim\limits_{0\to 1} v(x) = \boldsymbol{viability}$
- $\lim\limits_{1\to 100} s(y) = \boldsymbol{scale}$
- $v(x) \neq s(y)$
- **pilot** $v(x)$,
  if viable >1, get($)
  else, die trying
- **deploy** ($), solve $s(y)$

Funding to Impact Plan

1. K23 received a 29 impact score.
   – **Aim 1**. Design and validate a large language model (LLM) capable of extracting and summarizing PAD information from unstructured notes.
   – **Aim 2**. Co-design a minimally viable prototype for a “clinician in the loop” automated chart review (ACR) platform.
2. Plan for R21 to R01 pipeline.

Funding to Impact Plan

1. K23 recei

– **Aim 1.** D … cting and summar

– **Aim 2.** C … utomated chart re

2. Plan for R

Feasibility

R21
Backend system capa

R21
Front-end user desira

Department of Health and Human Services
National Institutes of Health
NATIONAL HEART, LUNG, AND BLOOD INSTITUTE

Notice of Award
FAIN# K23HL181388
Federal Award Date
08/18/2025

Recipient Information	Federal Award Information
1. Recipient Name TRUSTEES OF INDIANA UNIVERSITY 980 INDIANA AVE RM 2232 INDIANAPOLIS, IN 46202	**11. Award Number** 1K23HL181388-01
2. Congressional District of Recipient 07	**12. Unique Federal Award Identification Number (FAIN)** K23HL181388
3. Payment System Identifier (ID) 1356001673A1	**13. Statutory Authority** 42 USC 241 42 CFR 52
4. Employer Identification Number (EIN) 356001673	**14. Federal Award Project Title** An intelligent clinical decision support system for peripheral arterial disease
5. Data Universal Numbering System (DUNS) 603007902	**15. Assistance Listing Number** 93.837
6. Recipient's Unique Entity Identifier SHHBRBAPSM35	**16. Assistance Listing Program Title** Cardiovascular Diseases Research
7. Project Director or Principal Investigator Andrew Alexander Gonzalez, MD andrewg@iu.edu 312-259-7893	**17. Award Action Type** New Competing **18. Is the Award R&D?** Yes

Funding to Impact Plan

1. K23 funded(K23HL181388)
   – **Aim 1**. Design and validate a large language model (LLM) capable of extracting and summarizing PAD information from unstructured notes.
   – **Aim 2**. Co-design a minimally viable prototype for a “clinician in the loop” automated chart review (ACR) platform.
2. Plan for R21 to R01 pipeline.

Feasibility	**Impact**
R21 Backend system capability	
R21 Front-end user desirability	

Funding to Impact Plan

1. K23 funded(K23HL181388)
   - **Aim 1**. Design and validate a large language model (LLM) capable of extracting and summarizing PAD information from unstructured notes.
   - **Aim 2**. Co-design a minimally viable prototype for a “clinician in the loop” automated chart review (ACR) platform.
2. Plan for R21 to R01 pipeline.

Feasibility	**Impact**
R21 Backend system capability	**R01 Clinical Trial** ***Intervention***: Digital Health Intervention vs usual Care ***Outcomes***: Clinic throughput, safety, patient understanding, lost to follow-up
R21 Front-end user desirability	

Funding to Impact Plan

1. K23 funded(K23HL181388)
   - **Aim 1**. Design and validate a large language model (LLM) capable of extracting and summarizing PAD information from unstructured notes.
   - **Aim 2**. Co-design a minimally viable prototype for a “clinician in the loop” automated chart review (ACR) platform.
2. Plan for R21 to R01 pipeline.

Feasibility	**Impact**	
R21 Backend system capability **R21** Front-end user desirability	**R01 Clinical Trial** ***Intervention***: Digital Health Intervention vs usual Care ***Outcomes***: Clinic throughput, safety, patient understanding, lost to follow-up	**R01** • **[NLM]** Data curation at scale • Horizontal scaling (add other disease verticals) • Vertical scaling (add model capabilities)

Subproblems with different boundary conditions:

R21
Backend system capability

- $\lim_{0 \to 1} v(x) = \boldsymbol{viability}$
- $\lim_{1 \to 100} s(y) = \boldsymbol{scale}$
- $v(x) \neq s(y)$
- **pilot** $v(x)$,
  if viable >1, get($)
  else, die trying
- **deploy** ($), solve $s(y)$

Subproblems with <u>different boundary conditions</u>:

} $\lim\limits_{0\to 1} v(x) = \boldsymbol{viability}$

} $\lim\limits_{1\to 100} s(y) = \boldsymbol{scale}$

} $v(x) \neq s(y)$

} **pilot** $v(x)$,

if viable >1, get($)

else, die trying

} **deploy** ($), solve $s(y)$

R21
Backend system capability

R21
Front-end user desirability

Subproblems with <u>different boundary conditions</u>:

- $\lim\limits_{0\to1} v(x) = \boldsymbol{viability}$
- $\lim\limits_{1\to100} s(y) = \boldsymbol{scale}$
- $v(x) \neq s(y)$
- **pilot** $v(x)$,
  if viable >1, get($)
  else, die trying
- **deploy** ($), solve $s(y)$

R21
Backend system capability

R21
Front-end user desirability

R01 Clinical Trial
Intervention: Digital Health Intervention vs usual Care
Outcomes: Clinic throughput, safety, patient understanding, lost to follow-up

Subproblems with different boundary conditions:

- $\lim_{0 \to 1} v(x) = \boldsymbol{viability}$
- $\lim_{1 \to 100} s(y) = \boldsymbol{scale}$
- $v(x) \neq s(y)$
- **pilot** $v(x)$,

```
if viable >1, get($)
else, die trying
```

- **deploy** ($), solve $s(y)$

R21
Backend system capability

R01
- **[NLM]** Data curation at scale
- Horizontal scaling (add other disease verticals)
- Vertical scaling (add model capabilities)

R21
Front-end user desirability

R01 Clinical Trial
Intervention: Digital Health Intervention vs usual Care
Outcomes: Clinic throughput, safety, patient understanding, lost to follow-up

Implementing iCHAI trustworthiness framework for future grants and projects

Choice 1 - **ALCHEMI** Lab
Applied ***L***earning in ***C***linical ***H***ealthcare ***E***mpowered by ***M***achine ***I***ntelligence

Choice 2 - **ELACHI Lab**
Evidence-based ***L***earning and ***A***nalytics for ***C***linical ***H***ealthcare ***I***nsights

Implementing ~~iCHAI~~ trustworthiness framework for future grants and projects

iCHAI Trustworthiness Framework

Data	Model	Output	Implementation
	Mechanistic Interpretability		

iCHAI Trustworthiness Framework

Data	Model	Output	Implementation
Observability (of training data)	Clinical Interpretability	Accuracy (factually correctness)	Safety (does system cause harm)
Transparency (of ontology)	Mechanistic Interpretability	Alignment (clinical setting dependent)	Sustainability (w/respect to resources)
Provenance (of labels)	Validity (internal and external)	Verifiability (by humans or machines)	Reliability (across clinical scenarios)

Guiding Principles for Healthcare AI model development, pipeline deployment, and system implementation

1. System trustworthiness ≠ Trusted by humans
2. Trust is an essentially subjective human feeling… the ultimate human-on-the-loop safeguard.
   - **Why?:** Cannot be programed; unlikely to be an emergent behavior that can be described by a differential equation and optimized with a loss function
     - **Caveat**: Trust itself is an essentially human phenomenon, but machines can be programed to determine if a human would likely trust an output in a particular situation.
3. Trust is much easier to lose than to build.
4. Building trustworthy systems is as easy as its every going to be.
   - **Current state**: machine-human hybrid systems are (presumably) non-adversarial and tasked with solving relatively easy problems
   - **Future state**: Running out of training data….what if every agent isn't playing for the same team?

Transitioning to **multi-modal models** implemented in **multi-agent systems**

Hierarchical components of a <u>clinical</u> *socio-technical* system

1. {**Service**} (class: Clinical, research, administrative, legal, logistical, business)
2. Clinical {**activity**}
   - Purpose: *diagnosis*, *prognosis*, *treatment selection*
   - Task: <u>thinking</u>, <u>doing</u>, <u>recording</u>, <u>communicating</u>
3. Data processing {**pipeline**}
4. machine {**agents**}
5. Deep learning {**models**}

Example 1 – MV{*e*}
Image-text pair(s)

Initially, there was a short segment near-occlusive lesion of the proximal right AT,…

After treatment with a 4 mm x 6 mm Lutonix

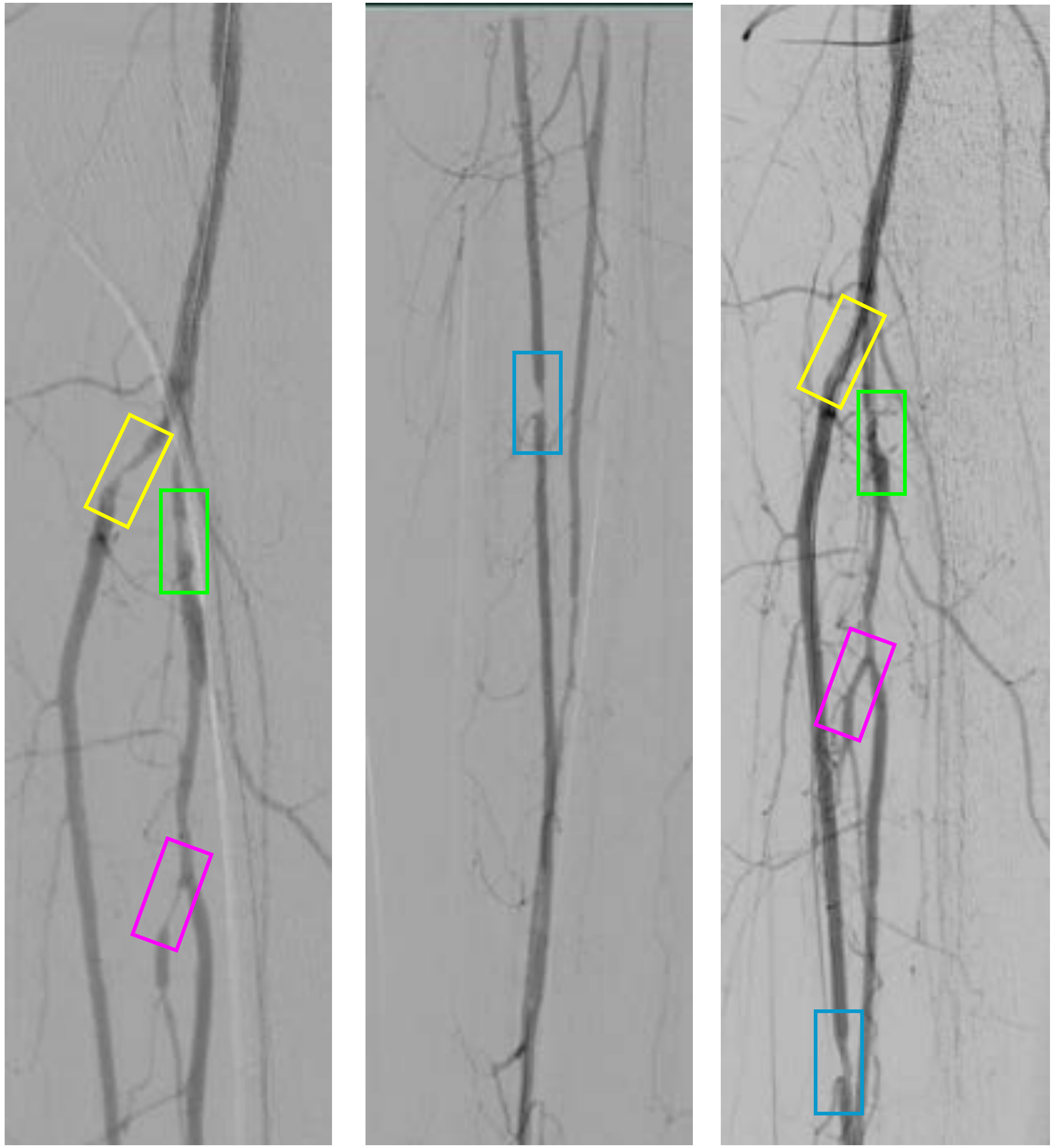
the proximal AT lesion resolved with no significant residual stenosis, the

Lesion	Lat.	Vessel	Location	State, pre-treatment	Treatment	State, post-treatment	Tx, Response	Image, pre	Image, post
PL-298	Right	Anterior tibial artery	Proximal 1/3	Near occlusive stenosis	4 mm Lutonix	Min residual stenosis	Adequate	Img_pl298_1	Img_pl298_2

Example 2 – Real world
Image-text pair(s)

short segment near-occlusive lesion of the proximal right AT short segment near occlusive lesion of the mid AT short segment moderate stenosis of the proximal right TP trunk a short segment near occlusive lesion of the right peroneal artery

the proximal AT lesion resolved with no significant residual stenosis mid-AT lesion improved but had a 30% residual stenosis relative to the adjacent normal artery the proximal TP trunk lesion had minimal residual stenosis but did not dilate up to the adjacent TP trunk artery proximal peroneal lesion resolved with no significant residual stenosis



Questions…?

andrewg@iu.edu

@dragonzMD (Insta)

@dragonzMD (twitter)

INDIANA UNIVERSITY
SCHOOL OF MEDICINE

50 years Regenstrief Institute

Supplementary Content

The role of Equity in Healthcare AI

Winning at equity with 4D Chess

1. Development
2. Data
3. Deployment
4. Decisions

Equity Considerations, a 4D Chess Approach

Set up	Level	Development	Data	Deployment	Decisions
Player	Individual / Institutional				
Opponent	Individual / Institutional				
The Board	System				

Considerations for model building

1. Data needed to build a model (type and amount)

2. What do you want the model to predict?

3. Output in a **useful form for end-users** in a **system that will listen** to those users (nudges and legitimacy; beware Semmelweiss)