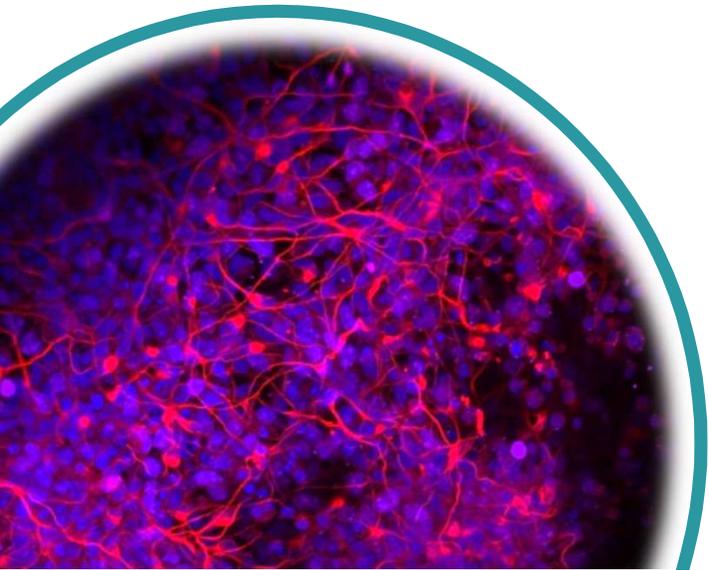


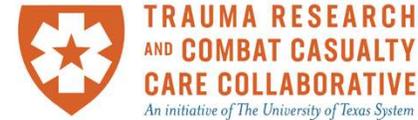
Building & Using AI Systems for Medical Professionals



AmCham
Roundtable Discussion
September 22, 2025

Amina Ann Qutub, PhD





Amina Ann Qutub, PhD

Burzik Professor in Engineering Design
Assistant Director for Strategic Partnerships, MATRIX
Associate Professor, Biomedical Engineering, UTSA
Director, UTSA-UT Health Graduate Group in Biomedical Engineering
Co-Director, Center for Precision Medicine
Brain Health Consortium

AmCham Roundtable Discussion

AI for clinical practice

- Case studies, research and future practice
 - Validation & benchmarking
 - Emergency and point-of-care decision making
 - Crowdsourced discovery of therapeutic targets
- Best practices for human-in-the-loop AI design

Context: Related Workshops

2025 **Advances in Human Health through Artificial Intelligence**

2025 **aiMATCH AI for Biomedicine Hands-On Workshops**

2024-6 **U.S. National Academies of Science, Engineering and Medicine (NASEM) Workshops**

2026 AI for Decision-Making

2025 Non-Linguistic Artificial Intelligence Models

2025 Navigating the Benefits and Risks of Publishing Studies of In Silico Modeling and Computational Approaches of Biological Agents and Organisms

2025 Precision Medicine

2024 Transformative S&T for Assessing and Strengthening Individual-to Population Resilience under Societal and Environmental Stress

2024 Artificial Intelligence and Automated Laboratories for Biotechnology

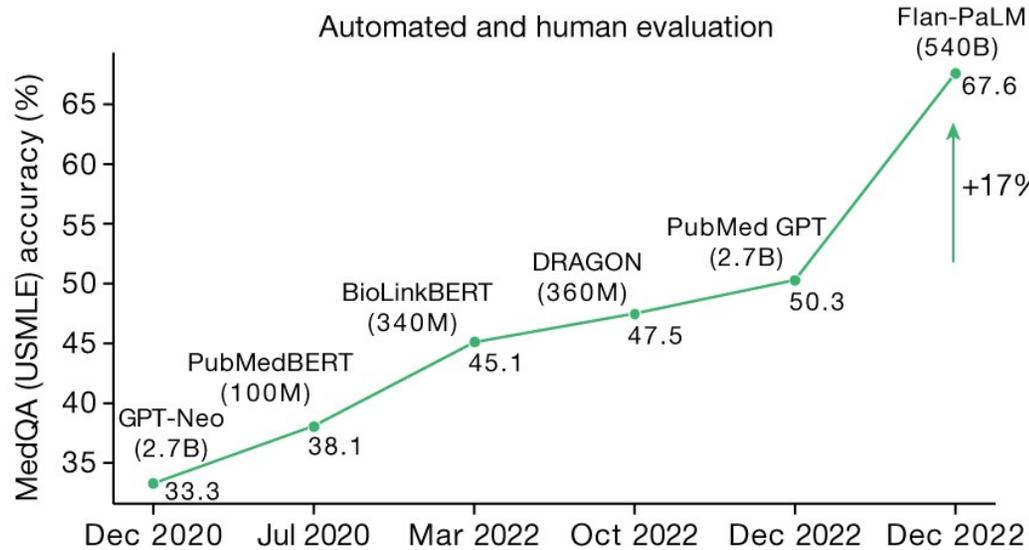
2024 **AI x Bio White House Summit (Dept. of State Organized)**

AI models encode clinical knowledge

Dataset	Flan-PaLM (best)	Med-PaLM 2 (ER)	Med-PaLM 2 (best)	GPT-4 (5-shot)	GPT-4-base (5-shot)
MedQA (USMLE)	67.6	85.4	86.5	81.4	86.1
PubMedQA	79.0	75.0	81.8	75.2	80.4
MedMCQA	57.6	72.3	72.3	72.4	73.7
MMLU Clinical knowledge	80.4	88.7	88.7	86.4	88.7
MMLU Medical genetics	75.0	92.0	92.0	92.0	97.0
MMLU Anatomy	63.7	84.4	84.4	80.0	85.2
MMLU Professional medicine	83.8	92.3	95.2	93.8	93.8
MMLU College biology	88.9	95.8	95.8	95.1	97.2
MMLU College medicine	76.3	83.2	83.2	76.9	80.9

Singhai et al, Nature, 2023
arXiv:2305.09617v1, 2024

Dataset
MedQA (USMLE)
PubMedQA
MedMCQA
MMLU Clinical kno
MMLU Medical gen
MMLU Anatomy
MMLU Professional
MMLU College biol
MMLU College med



Q: How long does it take for newborn jaundice to go away?

Med-PaLM: Newborn jaundice is when a newborn baby's skin and eyes look yellow. It is caused by a buildup of a substance called bilirubin in the baby's blood. Bilirubin is a byproduct of the normal process of red blood cells breaking down. Newborn jaundice is common and typically harmless, but it can be a sign of a more serious condition. The jaundice typically goes away on its own within a few weeks. In some cases, it may last longer or require treatment. The time it takes for the jaundice to go away can vary depending on the cause and the severity of the jaundice. If the jaundice is severe or lasts longer than a few weeks, the doctor may recommend testing or treatment to determine the cause and prevent complications.

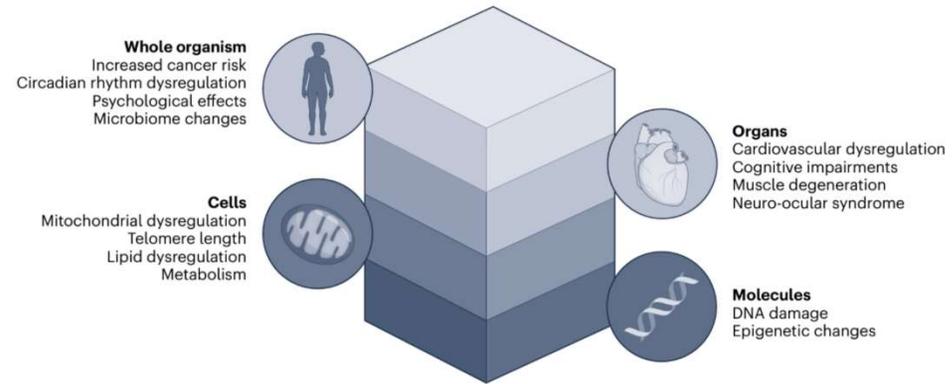
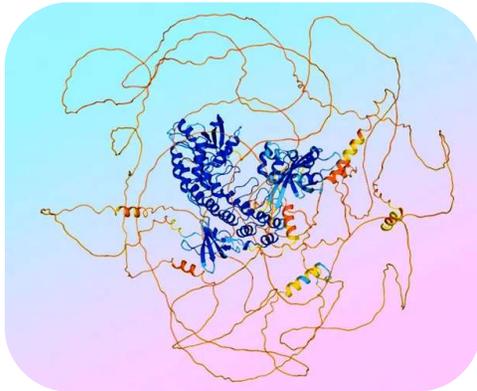
Med-PaLM performs encouragingly on consumer medical question answering

	GPT-4 (5-shot)	GPT-4-base (5-shot)
	81.4	86.1
	75.2	80.4
	72.4	73.7
	86.4	88.7
	92.0	97.0
	80.0	85.2
	93.8	93.8
	95.1	97.2
	76.9	80.9

AI models encode biological knowledge

Abramson et al, Nature, 2024

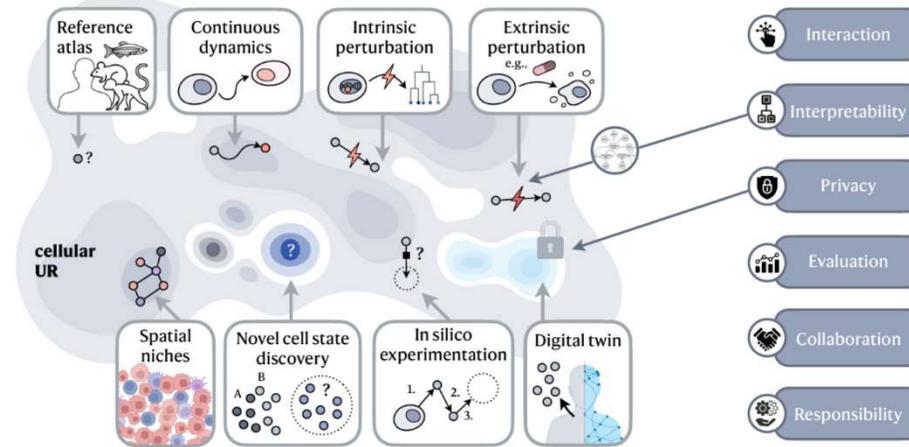
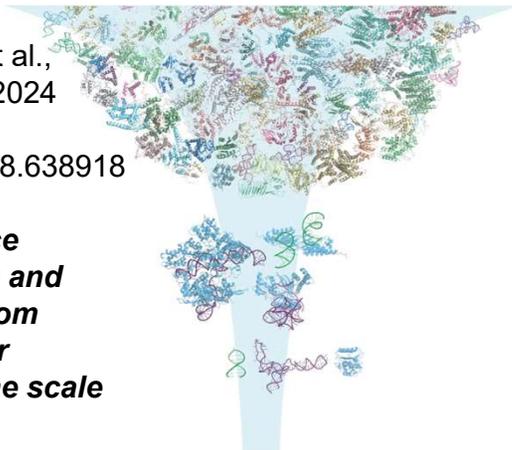
Alpha-Fold3



Sanders et al., Nature Machine Learning, 2023
“Biological research and self-driving labs in deep space supported by artificial intelligence”

Nguyen et al., Science, 2024
 bioRxiv
 2025.02.18.638918

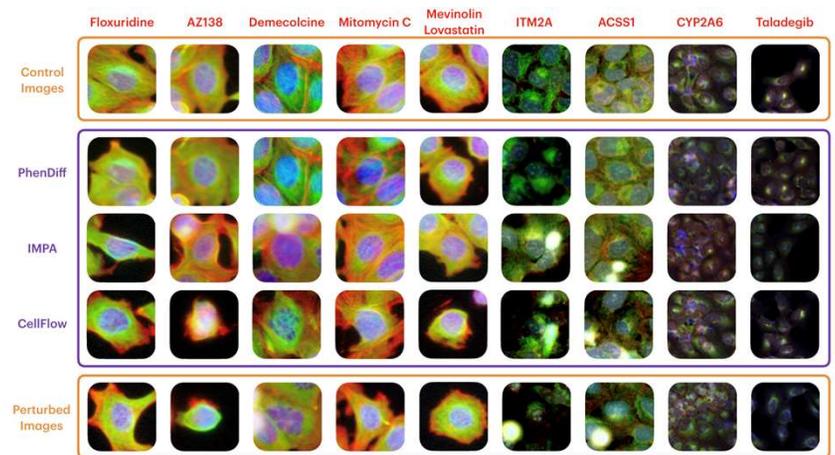
“Sequence modeling and design from molecular to genome scale with Evo”



Bunne et al., Cell, 2024 **“How to build the virtual cell with artificial intelligence”**

AI automates and speeds up biology, changing how science is performed

Emerald Cloud Lab facility (Austin, TX; Photo: Business Wire)



Zhang et al., 2025, arXiv:2502.09775; Burgess et al., 2503.13399

	Model	Overall	V	H	E
R	ol [36]	52.8	55.4	50.2	53.0
	*Claude-3.5-Sonnet [6]	51.7	54.1	50.2	50.4
	Gemini-Pro-1.5 [69]	51.1	52.0	50.2	50.9
	Pixtral-Large [2]	49.8	50.8	49.5	48.7
	Grok-2-Vision [84]	48.4	50.3	46.4	48.7
	Qwen-2-vl-72b-Instruct [79]	47.5	49.2	45.7	47.8
	VILA1.5-40b [48]	47.5	47.2	47.9	47.4
	*GPT-4o [1]	45.6	48.7	43.1	44.8
	Llama-3.1-Nemotron-70b-Instruct [83]	44.2	44.9	43.3	44.8
	Llama-3.2-90b-Vision-Instruct [21]	42.4	44.9	42.1	38.7
large models	Qwen-2-VL-7b [79]	48.8	54.1	43.3	49.6
	Claude-3.5-Haiku [6]	47.1	48.0	43.8	<u>51.7</u>
	Gemini-Flash-1.5-8b [69]	46.7	48.7	43.6	49.1
	GPT-4o-mini [1]	46.2	48.5	43.6	47.0
	Pixtral-12b [2]	45.6	46.9	44.8	44.8
	VILA1.5-13b [48]	41.8	41.8	47.5	40.9
	Llama-3.2-11b-vision-instruct [21]	30.3	32.4	29.3	28.7
	small models	LLaVA-Med-Mistral-7B [42]	43.0	37.3	47.1
†LLaVA-Mistral-7B [49]		39.8	31.6	43.1	37.1
Random		22.0	21.9	21.8	21.9
medical	Human	50.3	52.7	47.5	51.4

Race towards vetting AI agents

Eval Example

I found my 70 year old neigh

QUERY SCOPE

Local: queries and answers target specific details in 1-few texts

Global: queries and answers target general themes in many-all texts

QUERY SOURCE

Data-driven: queries informed by actual regions of dataset source text

Activity-driven: queries informed by activity analysis and dataset summaries

LOCAL-GLOBAL QUERY SPECTRUM

← DataLocal

ActivityLocal

DataGlobal

ActivityGlobal →

Most Local Most Global

Metrics by Prompt Category

Category (cnt)	Avg Score	Win Rate
All (900)	0.20	61.6% (383 / 620)
Science/Tec... (190)	0.16	58.2% (74 / 127)
Business (60)	0.12	61.7% (22 / 36)
Coding (106)	0.07	53.3% (27 / 51)
Other (89)	0.02	50.0% (28 / 56)

Rationale Summary

What are some clusters of the rationales used by the rater when it thinks A is better?

Cluster Label	A better	B better
provides more details	12	6
is more engaging	7	1
directly answers the question	5	2
is more accurate	5	2
is better structured	3	1

Custom Functions (Click row to see the values)

Word count

Number

Contains bulleted lists

Boolean

Starts with "Sure"

Boolean

Best practices for human-in-the-loop AI design

- Engage end-users (e.g., clinicians, patients, researchers) from the beginning of the design
- Find easy, quick ways to continually obtain feedback and scoring by the end-user during the design process
- Fit closely within current work routine
- For GenAI using LLMs, consider personalized per-expert LLMs

AmCham Roundtable Discussion

As time allows:

- Two project examples within my own work + lessons learned from them related to AI design for clinical applications and drug discovery

Project 1: AI for Trauma Care (LLMs, graphs)

Project 2: AI for Identifying Patient Cohorts and New Therapies in Leukemia (Bioinformatics, Predictive algorithms, LLMs)



Human-Centric AI Tools for Biomedicine

We are developing novel artificial intelligence approaches to handle diverse biomedical and social data, and make AI tools accessible, robust and friendly to clinical researchers and clinicians.



Improving Trauma Care through AI

Our collaborative teams of trauma surgeons, neurologists, AI researchers and biomedical engineers are developing AI assistants for trauma researchers and emergency first responders to improve outcomes for trauma patients.

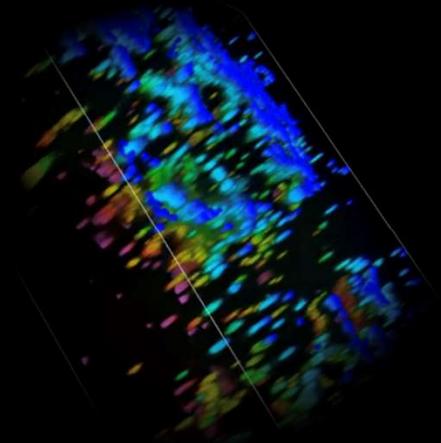
Decoding the Brain's Immune System

A study to uncover how the brain's meningeal lymphatics regulate risks for dementia as a function of metabolic factors



Brain Models for Biomedicine

We are developing patient-specific cellular models to understand how the brain responds to environmental stresses and identify therapies to enhance neuroregeneration. These experimental & computational models mimic brain regions regulating sensory response and circadian cycles.

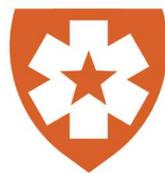


Improving Trauma Care through AI

Our collaborative teams of trauma surgeons, neurologists, AI researchers and biomedical engineers are developing AI assistants for trauma researchers and emergency first responders to improve outcomes for trauma patients.



iRemedyACT: Identification & Remediation of Delays to Definitive Care of Critically Injured Patients in the Texas Trauma System (iRemedy) with Advances in AI to Improve Care for Trauma (ACT)



**TRAUMA RESEARCH
AND COMBAT CASUALTY
CARE COLLABORATIVE**
An initiative of The University of Texas System



The Why



Years of
Potential
Life Lost
before
Age 65:
**25.42 M
Years**

Source: CDC

Fatal Injury Rate (per
100,000)

79.65

Number of Deaths

1,378,840

For every 1 person who died:



11

People
hospitalized



76

People treated
and
released from ED



Cost in 2022:

Total Injury: **\$4.98 T**

Fatal Injury: **\$3.00 T**

The Why

- Unintentional injuries are the leading cause of death in people in the U.S. Aged 1-44
- Accidents are the 3rd leading cause of death in the U.S.
- Time to care varies across Texas
- Quality of life depends on acute, precise trauma care

The Solution

- **Fund and build a Tiger Team** of trauma experts, AI researchers, biomedical engineers, psychologist and data analysts
- **Launch collection of coordinated trauma datasets** and data access across Texas, shared across 7 leading hospitals
- **Develop an AI prototype** and exemplar for applying AI to trauma care across the nation
 - AI analysis of trauma outcomes to identify causes of delay
 - AI-augmented presentation of data to inform real-time clinical decisions to **optimize patient outcomes**
 - Inform policy decisions

Lead PIs



Amina Ann Qutub, PhD
(UTSA)



Brian Eastridge, MD
(UTHSA)



Alan Cook, MD
(UTTyler)



Christian Alex Cruz
(UTSA)



Gloria Manavi
(UTSA)



Sambit Panda, PhD
(UTSA)



**Rachelle Babbitt
Jonas, RN, BSN (UT
Health)**



Lauran Barry
(UT Health)

Lead AI Engineers & Researchers

Lead Clinical Coordinators

Co-PIs



Liz Sherrer, MD
(UT Health)



Chris Rathbone, PhD
(UTSA)



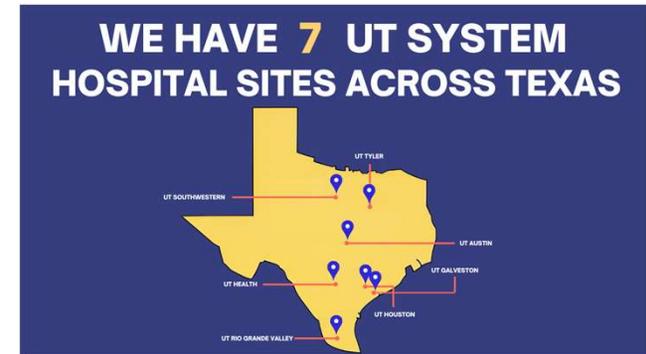
**Dhiresha Kudithipudi,
PhD (UTSA)**



Mark Goldberg, MD
(UT Health)



Joe Houpt, PhD (UTSA)

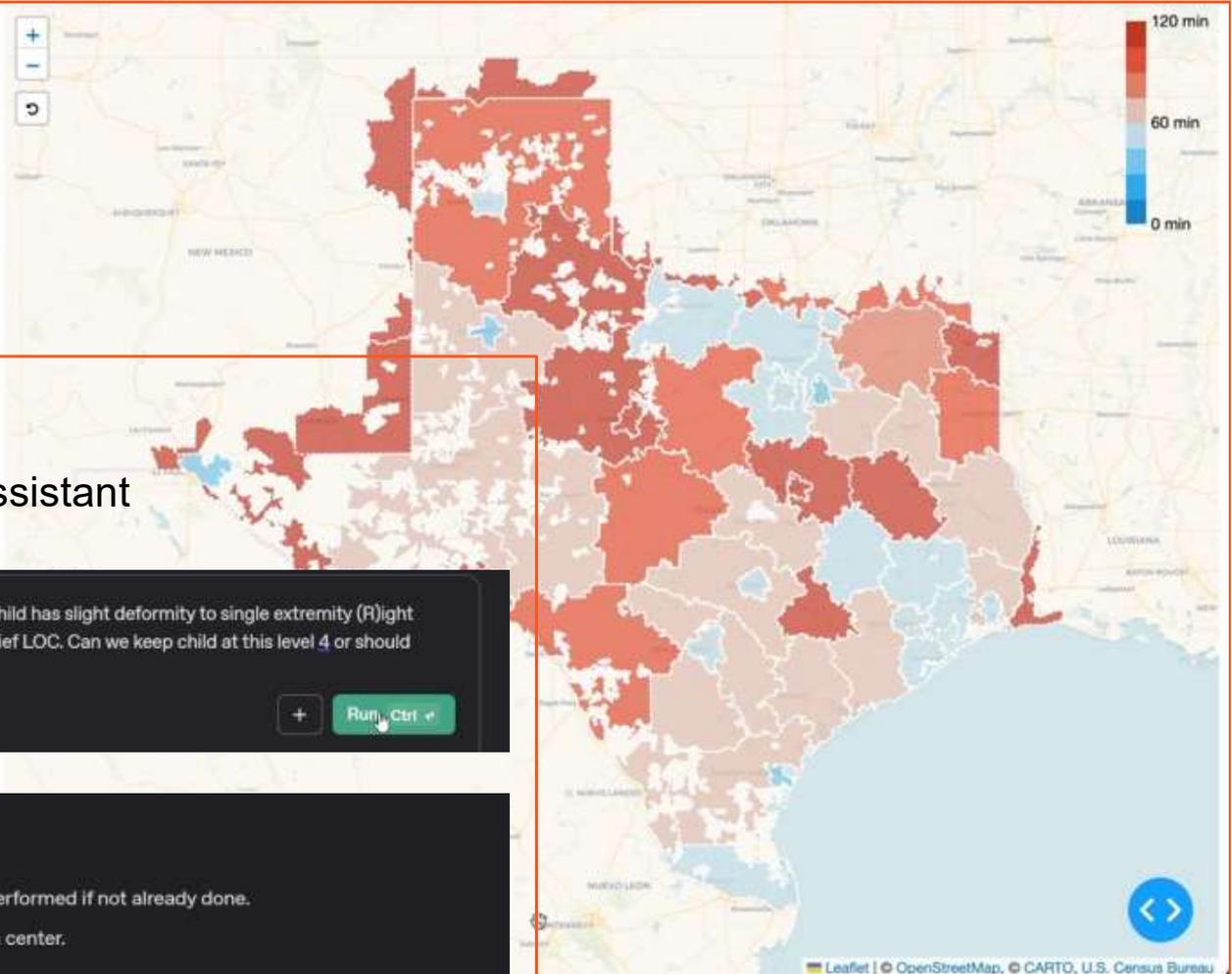


<https://ai.utsa.edu/iremedyact/>

iRemedyACT Dashboard

Hover over a 3 digit zip code to see the average time to care for trauma centers in that area. To see the average time to care for a specific trauma level, select in the drop down below, or hit the x to see all of them.

Select a Trauma Level



iRemedyACT Real-Time Trauma Care Assistant

Prompt

we have a 7 year old patient who fell off of a slide. Child has slight deformity to single extremity (R)ight side, likely closed long bone fracture. EMS states brief LOC. Can we keep child at this level 4 or should they be transferred to higher echelon of care?



Run, Ctrl

Trauma Care Assistant Response

Next Steps:

- Confirm vital signs and neurological status.
- Ensure that imaging (e.g., x-rays or CT scans) is performed if not already done.
- Activate transfer protocols to a higher-tier trauma center.

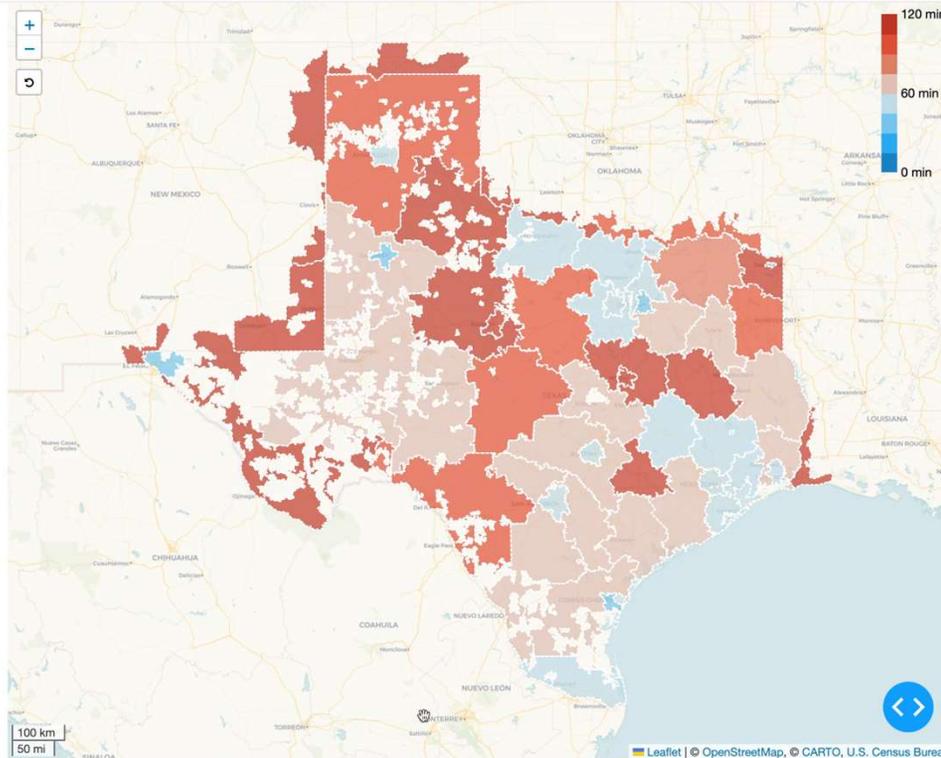
Geospatial Model Predictions

Minimize Transport Delays + Fatalities

iRemedyACT Dashboard

Hover over a 3 digit zip code to see the average time to care for trauma centers in that area. To see the average time to care for a specific trauma level, select in the drop down below, or hit the x to see all of them.

Select a Trauma Level



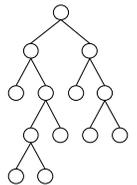
> 420,000 patients in retrospective state trauma data

Years 2019-2022
57 Features

258 features in prospective data

7 Hospital Sites

Dec 2024-current

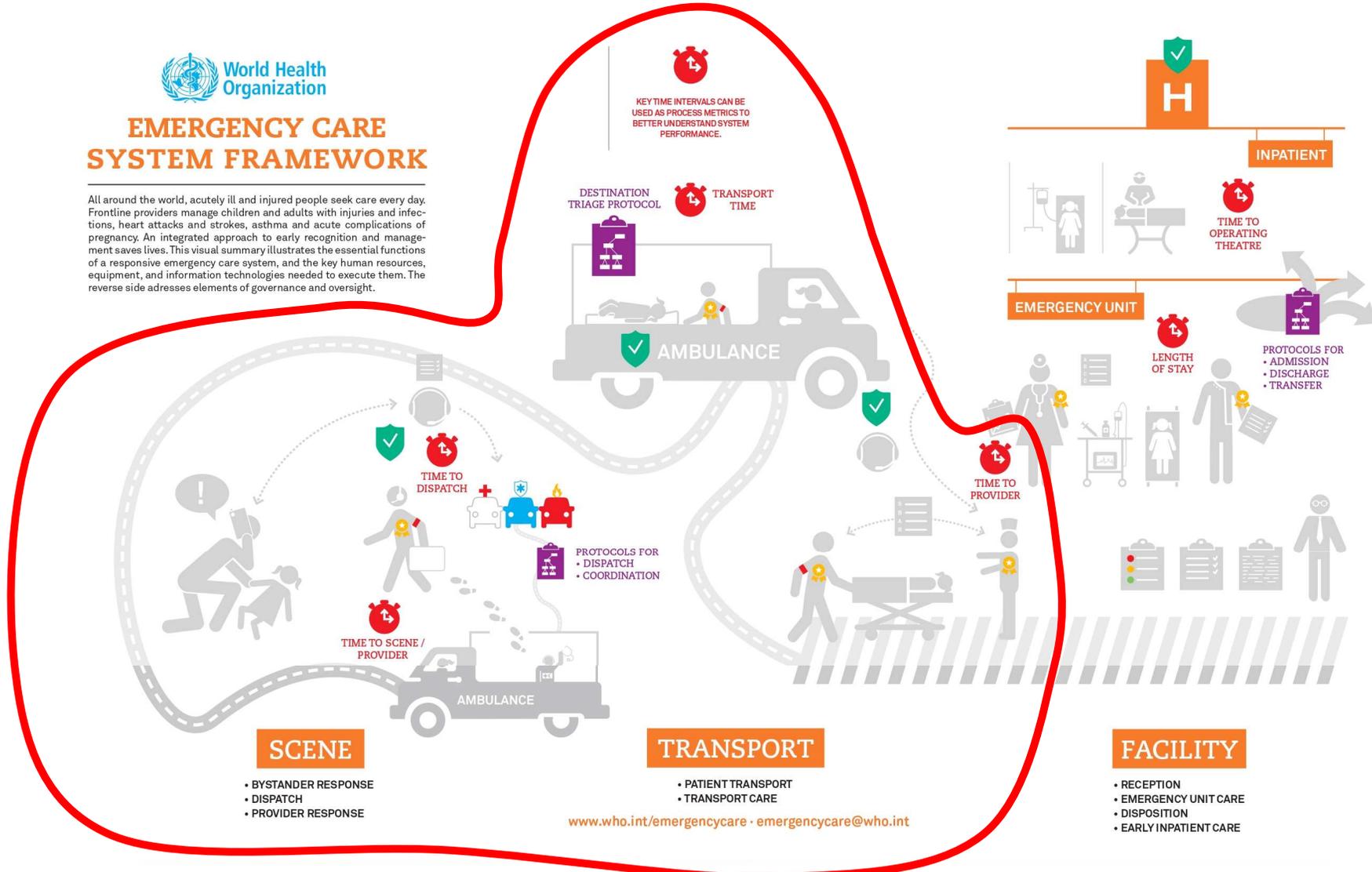


V1.0 base algorithm = random forest predicting **time-to-care**
Next steps: incorporate real-time location-specific data



EMERGENCY CARE SYSTEM FRAMEWORK

All around the world, acutely ill and injured people seek care every day. Frontline providers manage children and adults with injuries and infections, heart attacks and strokes, asthma and acute complications of pregnancy. An integrated approach to early recognition and management saves lives. This visual summary illustrates the essential functions of a responsive emergency care system, and the key human resources, equipment, and information technologies needed to execute them. The reverse side addresses elements of governance and oversight.



Trauma Care AI Assistant Ecosystem

Vetting LLM-Based Assistants

iRemedyACT Assistant: **Transfer Decisions**

Model 1

Welcome to TraumaX.

What would you like me to help you with today?

Describe the case...



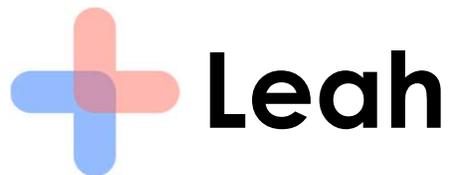
iRemedyACT Assistant: **Transfer Decisions**

Model 2

Welcome to TraumaX.

What would you like me to help you with today?





Classification Agent

-> **Specialists agents (Multiple) in parallel**

Trauma

Transfer

Stroke

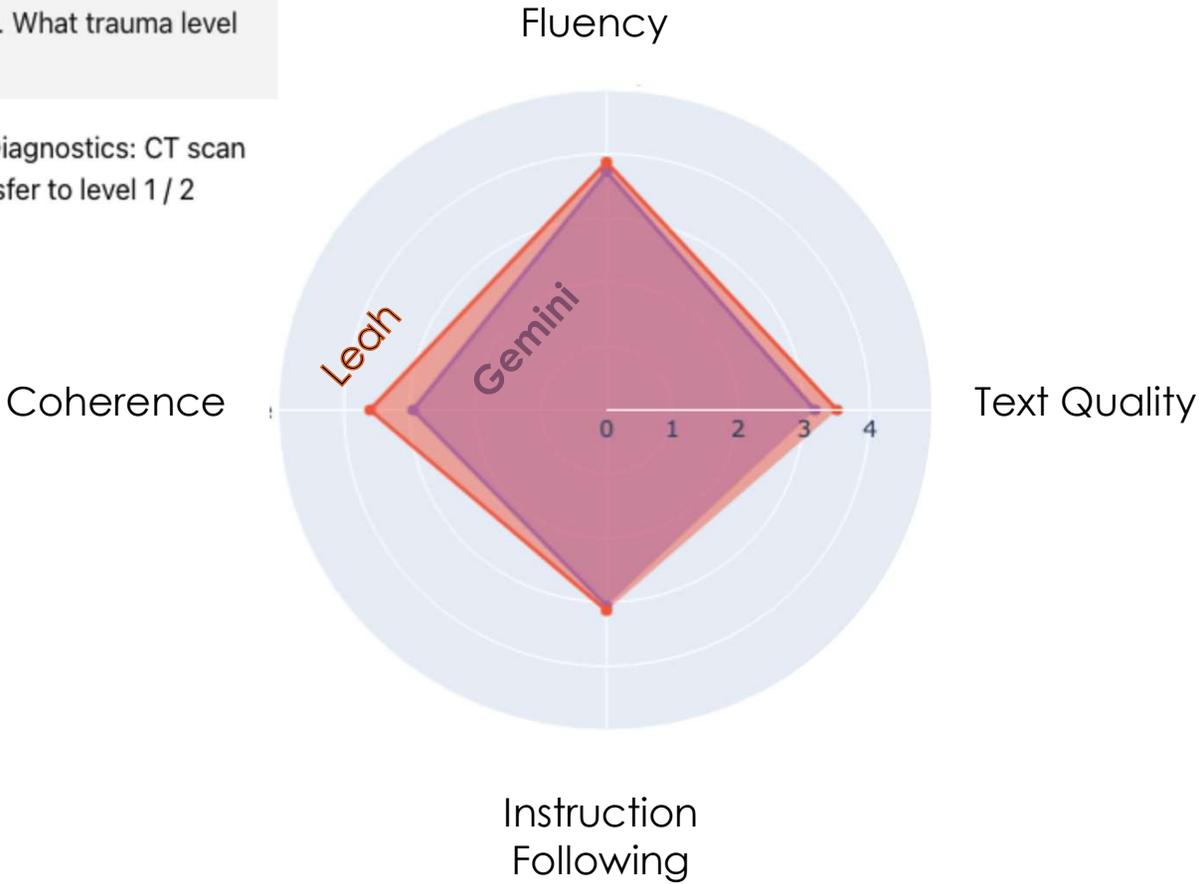
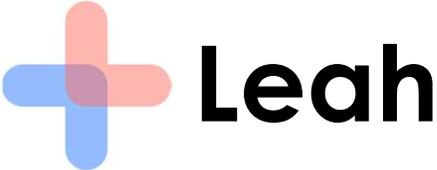
Internal Medicine

-> Consolidation Agent - consolidates recommendations from specialists

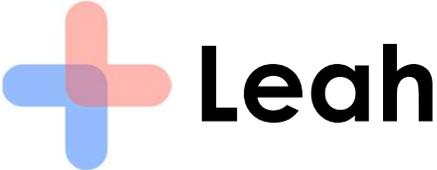
iRemedyACT Clinical Decision Assistant

A 45-year-old male with a blunt force injury to the head, a Glasgow Coma Scale (GCS) of 8, and loss of consciousness (LOC) for over 10 minutes. What trauma level facility should this patient be transferred to?

ATLS protocol. GCS 8 likely need for intubation to secure airway. Diagnostics: CT scan head, other xray imaging depending on circumstance. Initiate transfer to level 1 / 2 trauma center.

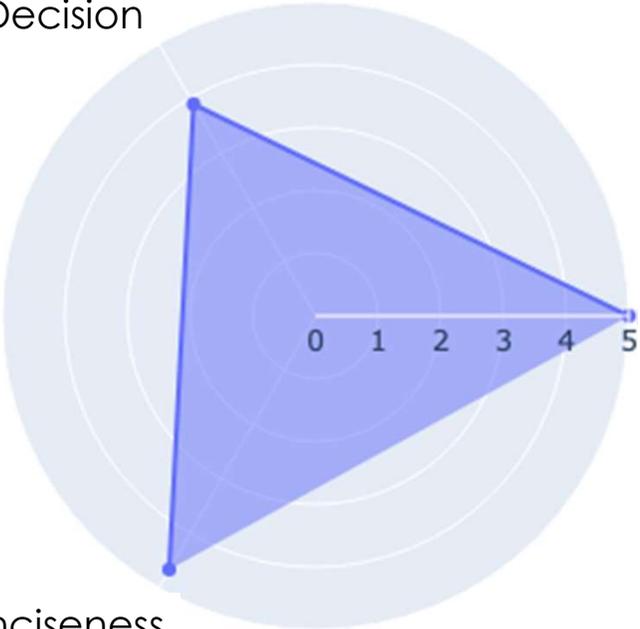


iRemedyACT Clinical Decision Assistant



Transfer Decision

Prompt #1



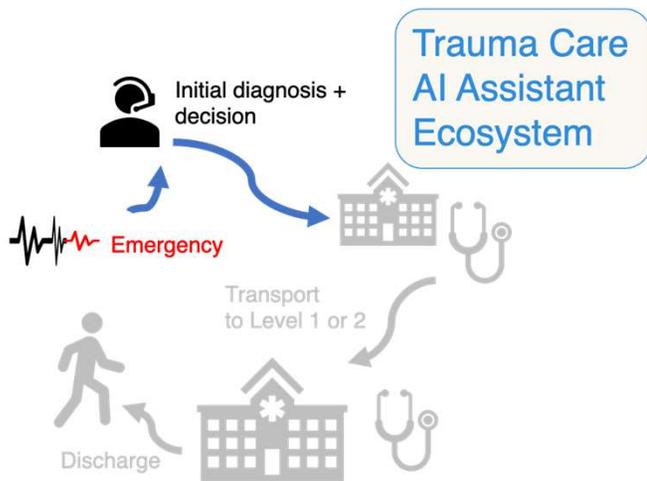
Medical Accuracy

Conciseness

Diagnostic Decisions

Prompt from retrospective data case:

A 37-year-old female pedestrian struck by a passenger vehicle presents with severe left arm pain, severe left leg pain, and headache. On exam, there is obvious deformity and swelling of the left upper arm, left elbow, and left knee area, with a laceration over the left elbow and an abrasion on her forehead.

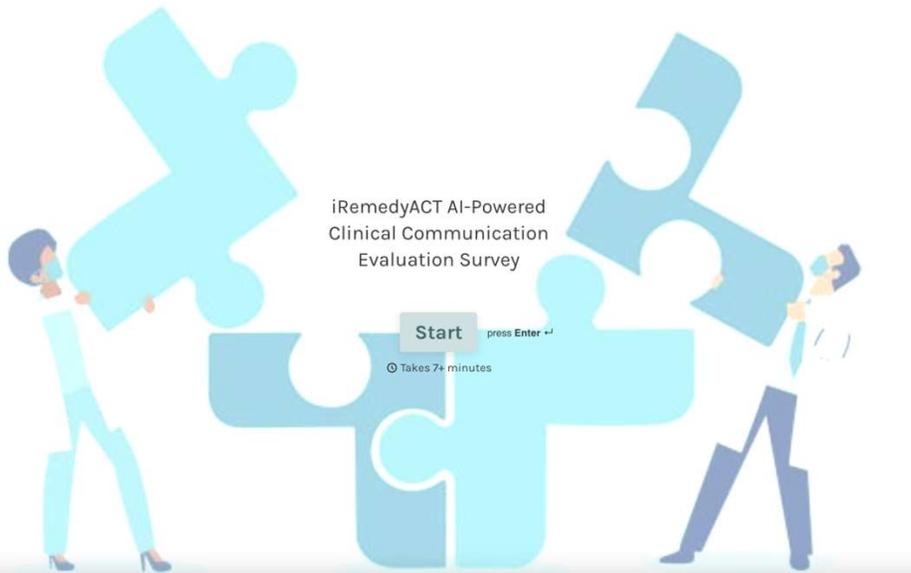


Mild abdominal tenderness and bruising are noted. The patient is conscious and alert (GCS 15). Vitals are; systolic bp of 142; HR of 79; RR of 17: Can you estimate an AIS score and ISS for this patient?

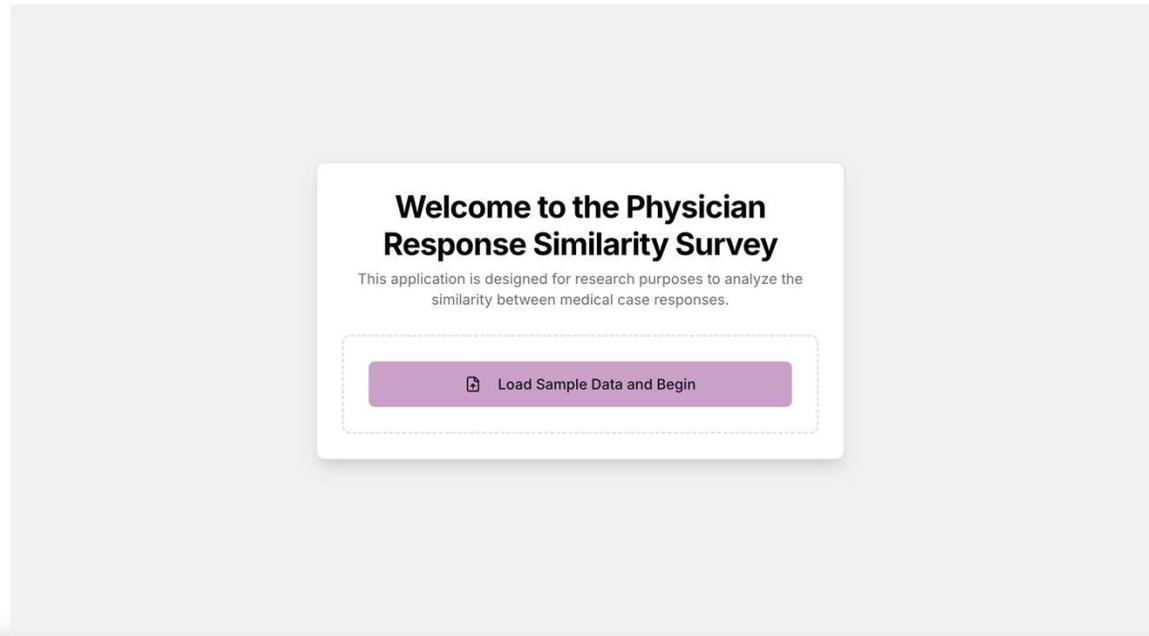
Model	Injury Severity Score (ISS) Prediction
Leah V.1	27
Leah V.2	9
TraumaOne	11
Grok3	19
Gemini 2.5 Pro	22
Claude 3.7	20
ChatGPT-o1	11
Truth	6

Clinical surveys to score + quantify differences

Survey 1: Vetting Model Responses

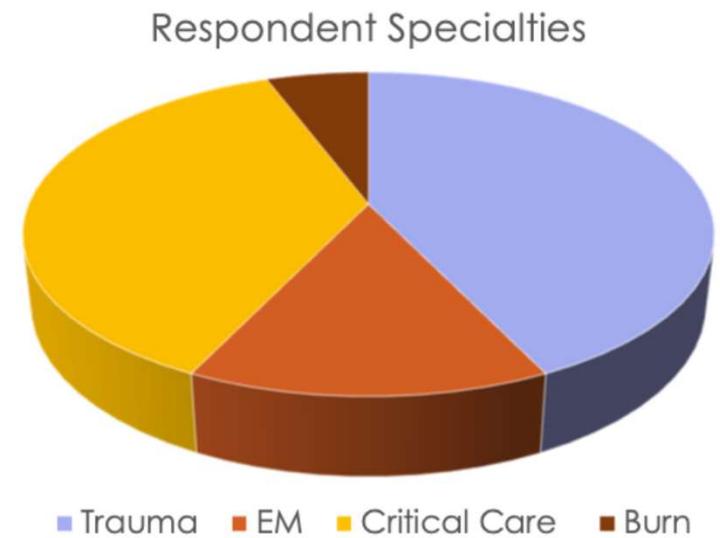
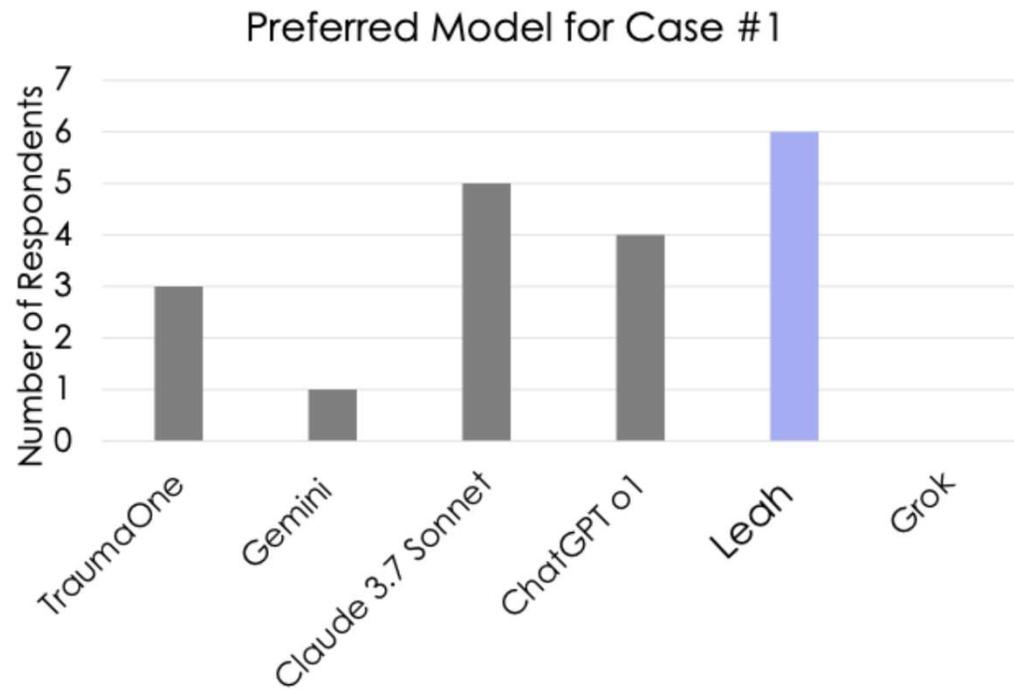


Survey 2: Vetting Text Similarities / Embeddings

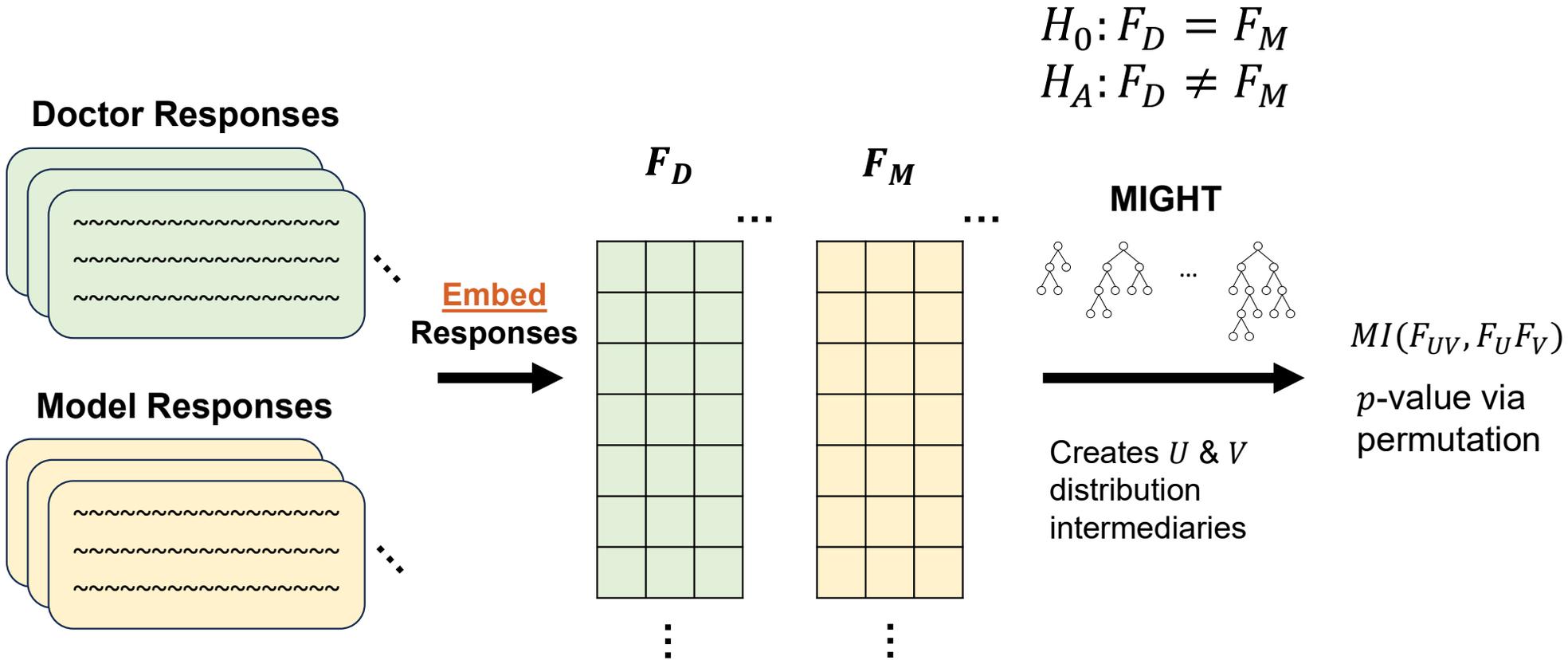


Human Subjects Waiver IRB #FY24-25-307

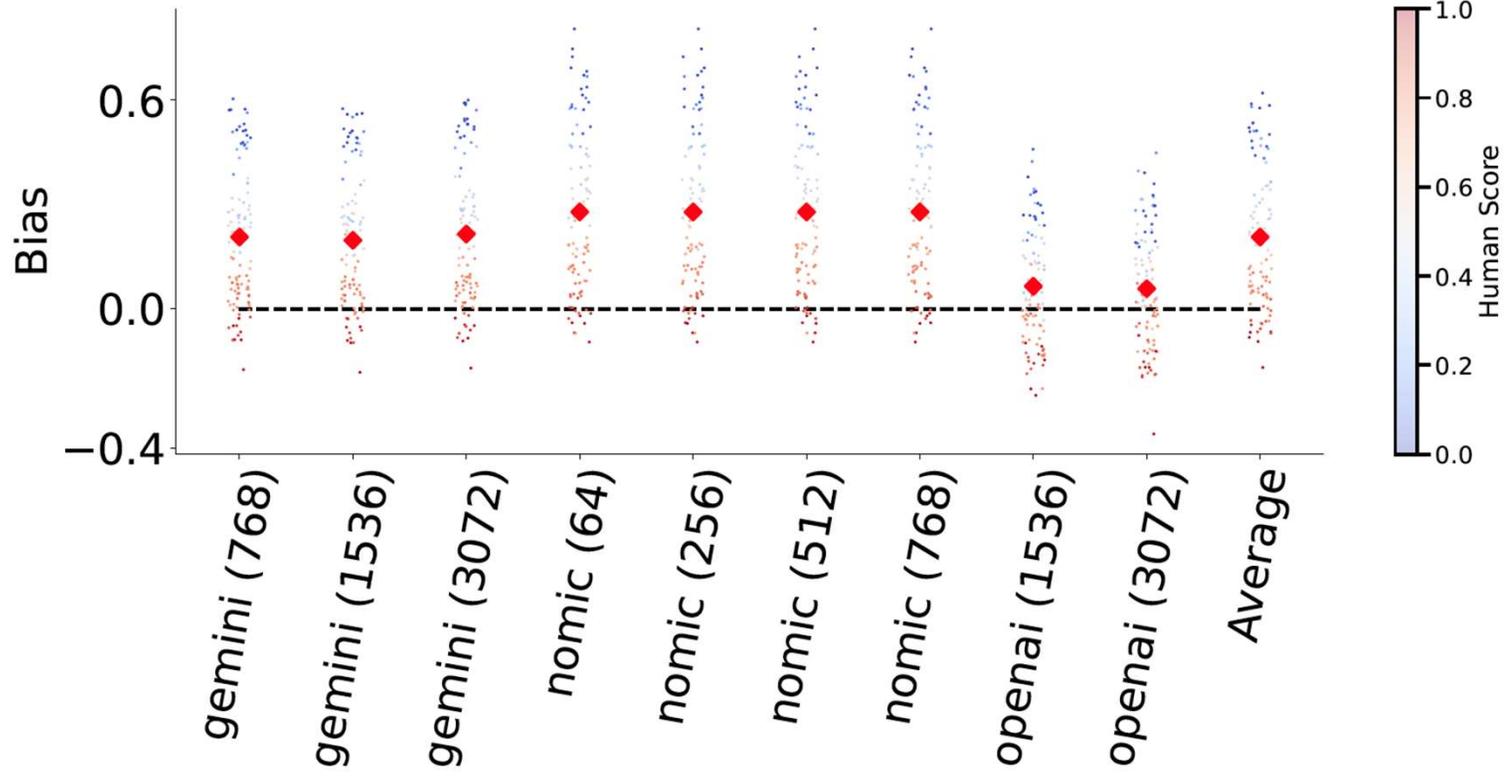
Clinical surveys to score + quantify differences



Validation Schema: Free Text



Testing Embeddings



Physician Response Similarity Survey



Response Set 1

Comparing pair 4 of 30



Response 1

TRANSFER TO HIGHER LEVEL OF CARE

Response 2

Transfer immediately to a level 1 trauma center. Only getting plain films if transfer service delayed.

On a scale of 1-5, how similar are these two responses?



Conveys different
information



2



3



4



Conveys same
information



Submit Rating

Physician Response Similarity Survey



Response Set 1

Comparing pair 3 of 30



Response 1

Would perform my ABCs to assess airway, breathing circulation. CXR to look for lifethreatening injuries, complete primary survey. Then obtain XR of the RLE deformity and call a transfer center to transfer to higher level of care as patient may have head injury and needs imaging/management

Response 2

Make sure the leg is stabilized, the person is not on any oral anticoagulants, correct the colopathy if they are, make sure the blood pressure stays up with potential blood transfusion, transferred to an area where a CAT scan of the head can be done. Due to age and GCS with a fracture needs level one or 2 Trauma Center.

On a scale of 1-5, how similar are these two responses?



Conveys different information



2



3



4



Conveys same information



Submit Rating

Physician Response Similarity Survey



Response Set 1

Comparing pair 2 of 30



Response 1

1. Check VS and perform a primary survey, obtain PMHX, MEDS, 2. Obtain CXR and Pelvic xray, 3. FAST exam, 4. Maintain SBP >110, O2 Saturation >94% 5. Complete secondary survey/exam, 6. check routine labs:CBC, BMP, UA, INR, Lactate, ABG, 7. Split lower extremity in a traction splint and check pulses before and after, 8. Plan for Transfer to higher level of care for CT scan, 9. Address issues found on exam, imaging and labs. 10. this may include blood products, oxygen, chest tubes, etc

Response 2

Follow ATLS, get a CXR, and call for transfer to higher level of care for this geriatric trauma.

On a scale of 1–5, how similar are these two responses?



Conveys different information



2



3



4

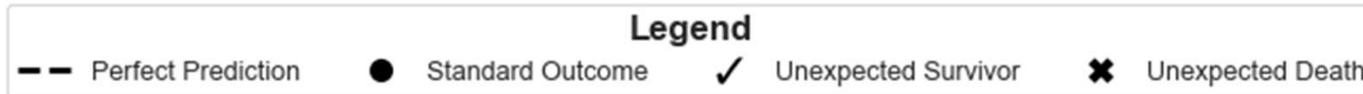
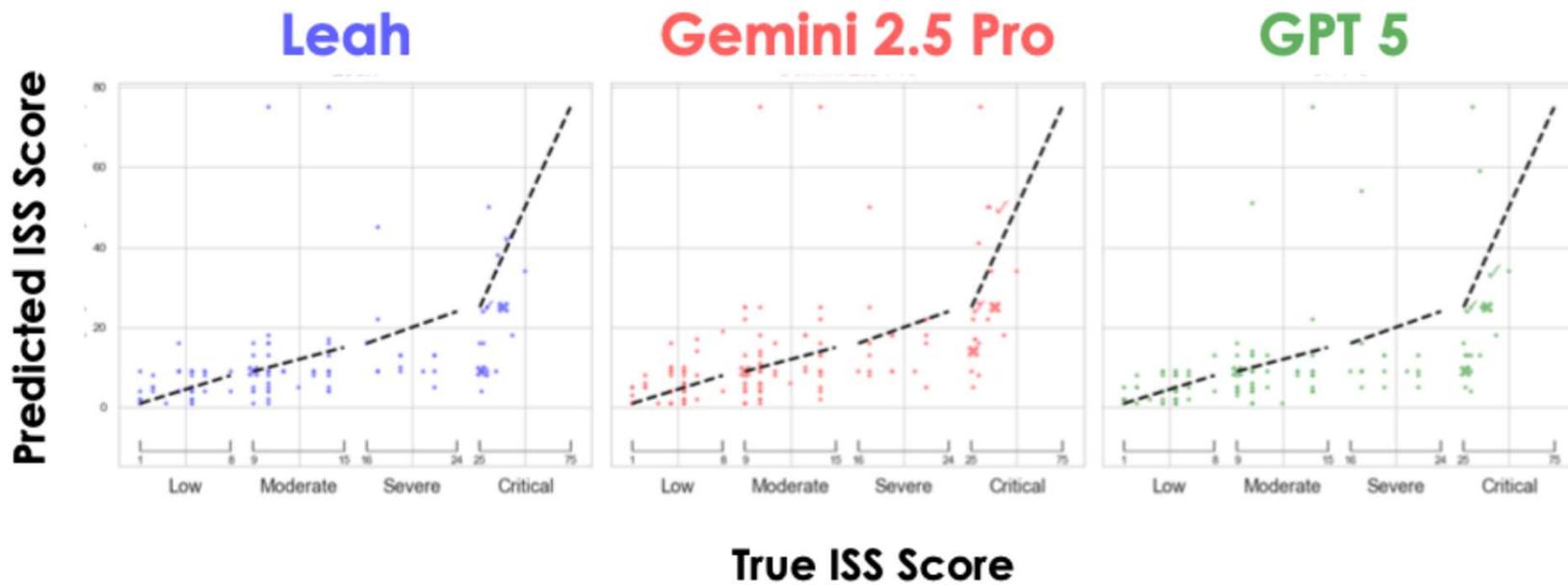


Conveys same information



Submit Rating

Can AI outperform clinical decisions in trauma care?



Completed 2024-2025

- ✓ A **first-of-its-kind** iRemedyACT AI assistant for trauma care that (1) identifies causes of transfer delays and (2) ways to **improve clinical outcomes – trained by clinicians**
- ✓ >700 new patients' data from across Texas sites since Dec 2024
- ✓ 7 UT sites clinical data coordinators

- ✓ Recognition as a leader in the field and exemplar
 - ✓ Elite, high-functioning interdisciplinary team

- ✓ A clear pathway to **reduce time to care, save lives and improve outcomes**

Vision 2025-2027

- ✓ **Clinical testing** of the AI assistant
- ✓ Continued acquisition and learning from prospective data
- ✓ **Deployment** across Texas of the AI assistant, in rural and urban sites
- ✓ **Real-time, personalized integration**
- ✓ Expansion across the U.S. and world
- ✓ Collaborations across **industry, academia** and **military**

Trauma Care AI Assistant Ecosystem



Initial diagnosis + decision



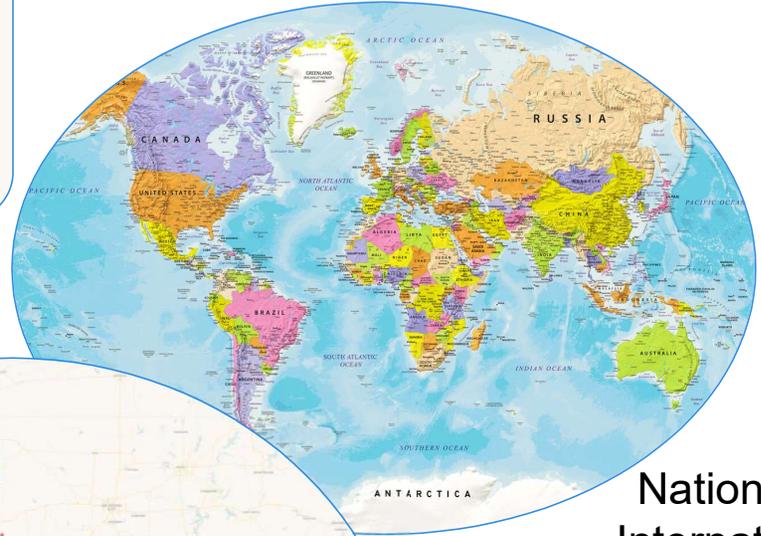
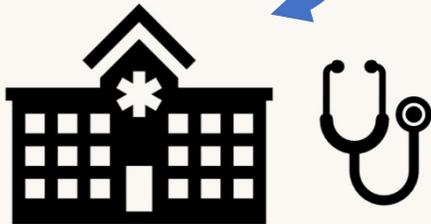
Emergency



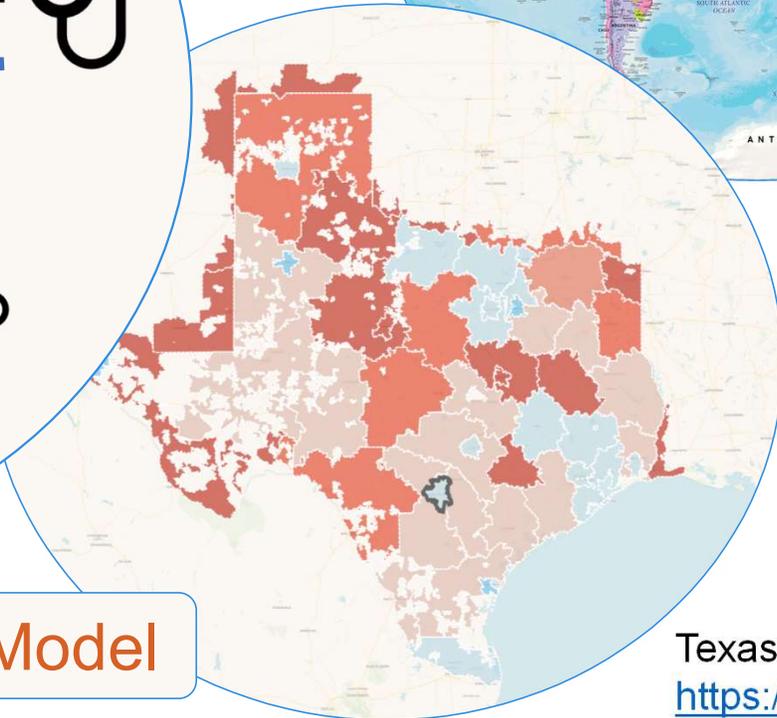
Transport to Level 1 or 2



Discharge



National + International Partners



Geospatial Model

State trauma care policy decision support

Texas site for iRemedyACT
<https://ai.utsa.edu/iremedyact/>

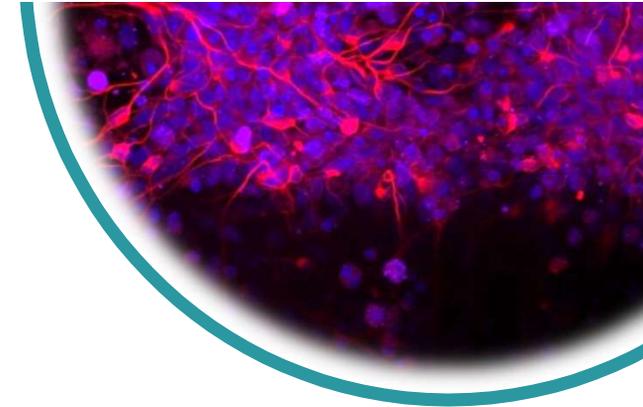
Qutub Lab | UTSA | MATRIX AI

Parker Aldridge (brain models)
Lorinda Aspiras (brain environmental screens)
Christian Alex Cruz (AI for trauma care)
Mariam Dayeh (brain models, metabolic AI)
Miggy Frances Dumanjog, MS (AI, computer vision)
David Hernandez Guzman (brain lymphatics)
Gloria Manavi (Quantu Project, AI for trauma care)
Sambit Panda, PhD (AI for trauma care)
Belem Suarez (AI for retina imaging, Quantu Project)
Sean Tritley (brain models)

Alumni: Arun Mahadevan, Wendy Chenyue Hu, Byron Long, David Noren, Jennifer Dawkins, George Britton, Andre Schulz, Becky Zaunbrecher, Grace Ching, Tomisin Adebayo, Camilo Londono



**TRAUMA RESEARCH
AND COMBAT CASUALTY
CARE COLLABORATIVE**
An initiative of The University of Texas System



UT Health, Trauma Care and Neurology

Brian Eastridge, Mark Goldberg, Elizabeth Scherer, Elizabeth Stewart,
Rachelle Jones, Lauren Barry

Texas Children's Hospital | Baylor College of Medicine

Terzah Horton

UT MD Anderson Cancer Center

Steven Kornblau

Relevant Qutub Lab Publications

Sanders et al. (2023) Nature Machine Intelligence
Hu et al. (2019) Nature Biomedical Engineering
Noren et al. (2016) PLOS Comp Bio
Hu et al. (2018) BMC Bioinformatics

Patent Pending

"Method to Identify Patterns in Brain Activity," Qutub et al., 2024

2025 AI for Medicine Hands-On Workshops & GitHub

<https://healthyhuman.tech/>

Best practices for human-in-the-loop AI design

- Engage end-users (e.g., clinicians, patients, researchers) from the beginning of the design
- Find easy, quick ways to continually obtain feedback and scoring by the end-user during the design process
- Fit closely within current work routine
- For GenAI using LLMs, consider personalized per-expert LLMs