

ARTIFICIAL INTELLIGENCE

# Artificial vs. Augmented Intelligence (AI) in Rheumatology:

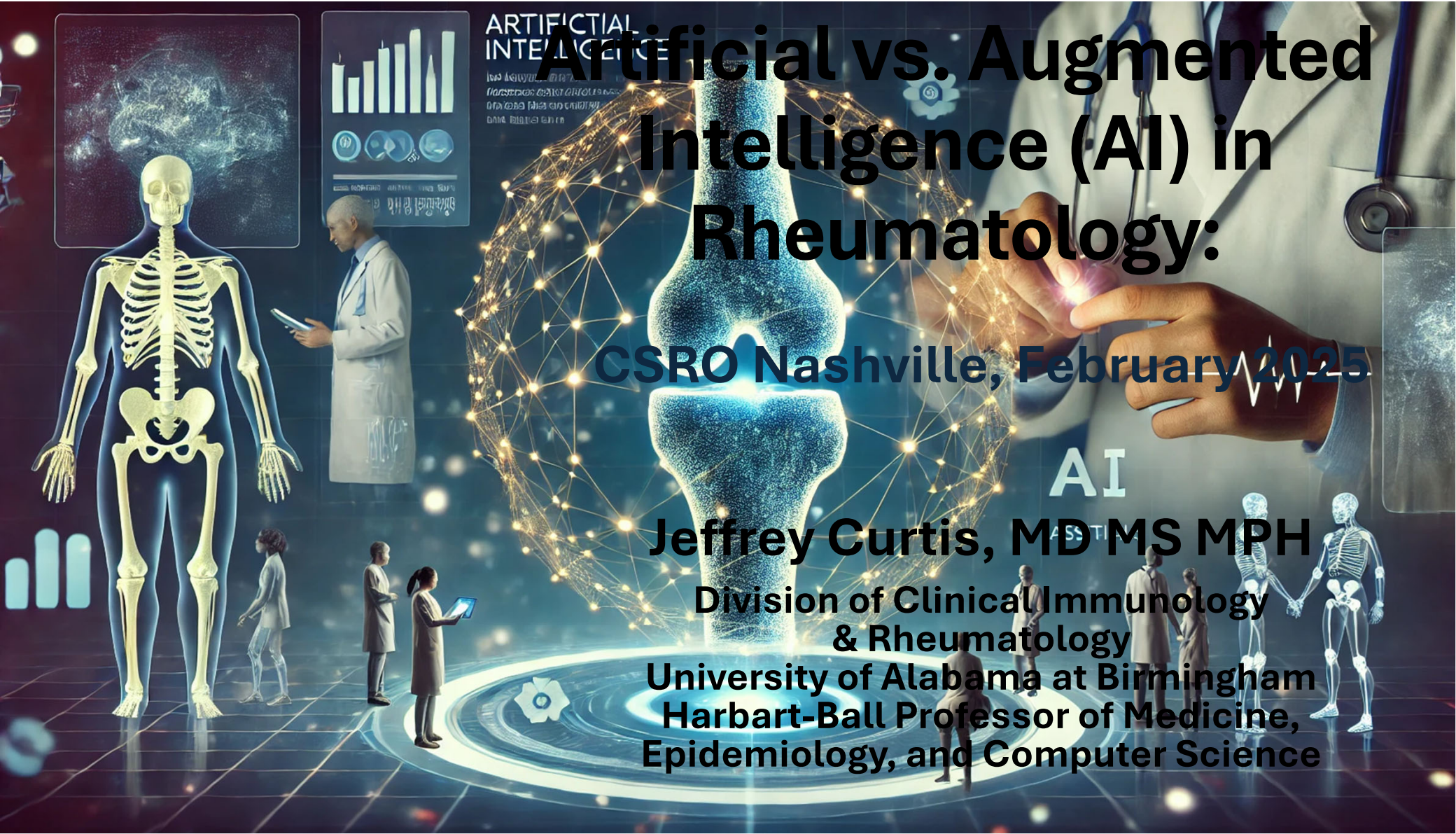
CSRO Nashville, February 2025

AI

Jeffrey Curtis, MD MS MPH

Division of Clinical Immunology & Rheumatology

University of Alabama at Birmingham  
Harbart-Ball Professor of Medicine,  
Epidemiology, and Computer Science



# Disclosures

	Company/Organization
<b>Honorarium / Consulting</b>	Abbvie, Amgen, AQTUAL, BMS, GSK, Janssen, Lilly, Novartis, Pfizer, Sanofi, Scipher, Sensimetric, Setpoint, UCB
<b>Research Grants</b>	Abbvie, Amgen, AQTUAL, BMS, GSK, Janssen, Lilly, Novartis, Pfizer, Sanofi, Scipher, Setpoint, UCB

**I lead several data coordinating centers including**  
**ACR RISE Registry**  
**PatientSpot patient registry (formerly ArthritisPower)**  
**Excellence Network in Rheumatology to Innovate Care and High-impact research (ENRICH) practice-based network**

# AI for Rheumatology

- What is AI and Where Did It Come From?
- Can AI help me be smarter (for my patients)?
- Enable me to find specific types of Patients
- Symptom checker, Flare Prediction -> Triage Patients
- Communicate with patients (e.g. chatbots)
- Read imaging to detect damage & inflammation
- Classify phenotypes & predict comorbidities
- Extracting disease activity data
- Predict treatment response to a new RA therapy
- Deliver on Value-Based Care



# The “Cost” of Technology



- The last time that doctors sat out on being at the table for a major technology implementation in medicine, we got the EHR
- Want to repeat that again?



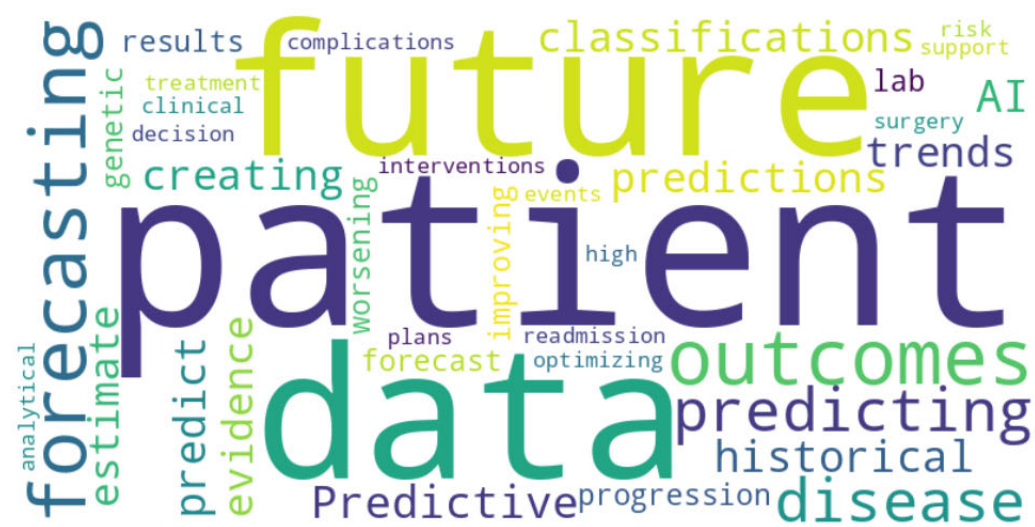
# Many Kinds of AI...

## AI on **Generative** vs. **Predictive** AI

Generative AI



Predictive AI



- Other types: Reactive, Limited Memory AI, Reinforcement Learning AI...
- Caution: some “AI” uses methods we’ve had, repackaged with a cool new name
- Much of predictive AI is simply a fancy way to create an equation based on data

# How Does Generative AI Work?

- **Consume much of the world's online content, literature, facts**
- **Deconstruct it into 'patterns' of content that commonly appear together**
- **Build an user interface that can predict content based on already-consumed context**
  
- **Prompt pattern: 'unlock' computations based on their context & structure**
  - **E.g. letter, scientific manuscript, clinical trial template**
  
- **Mental model: you send a text message to a 'friend'**
  - **Prompt: message that you send to AI**
  - **Output: response from AI**

**Finish this sentence, "You can't handle..."**

... **"the truth"** (no context)

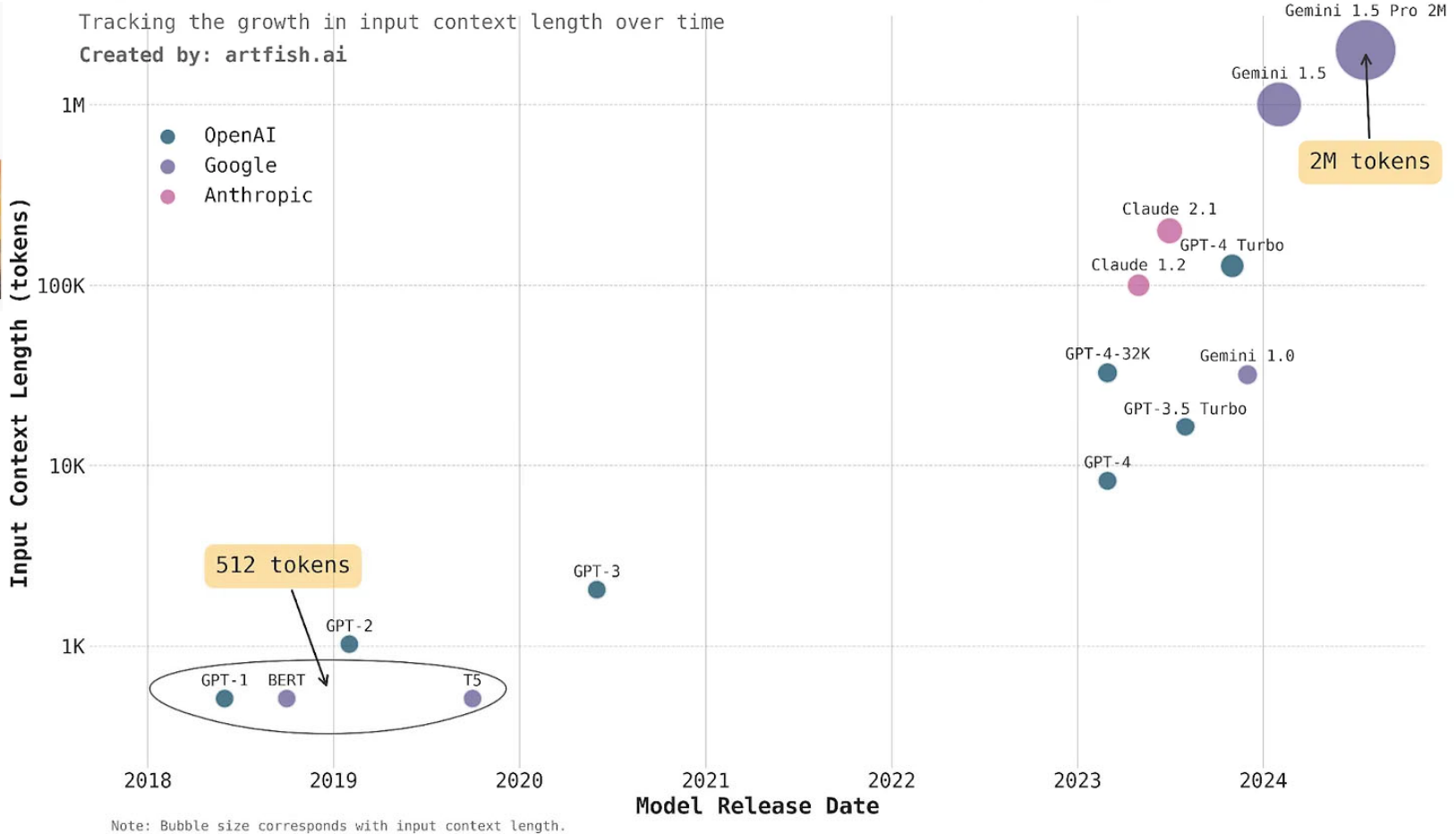
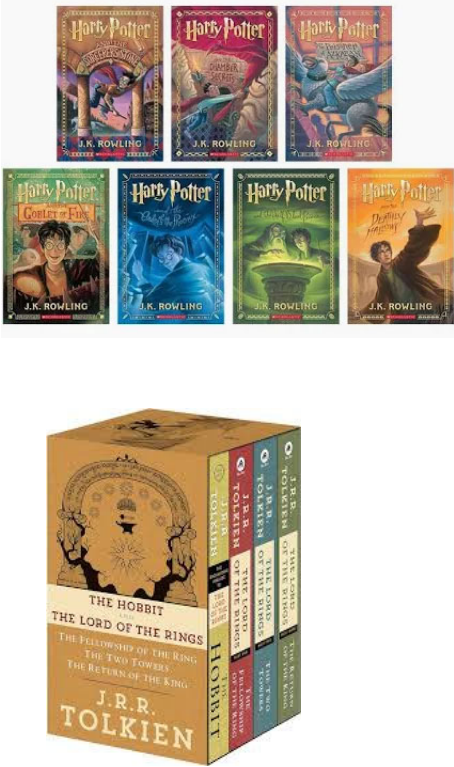
... **"all the fire selfies I'm about to drop!"**  
(persona of my 19yo daughter on Instagram)



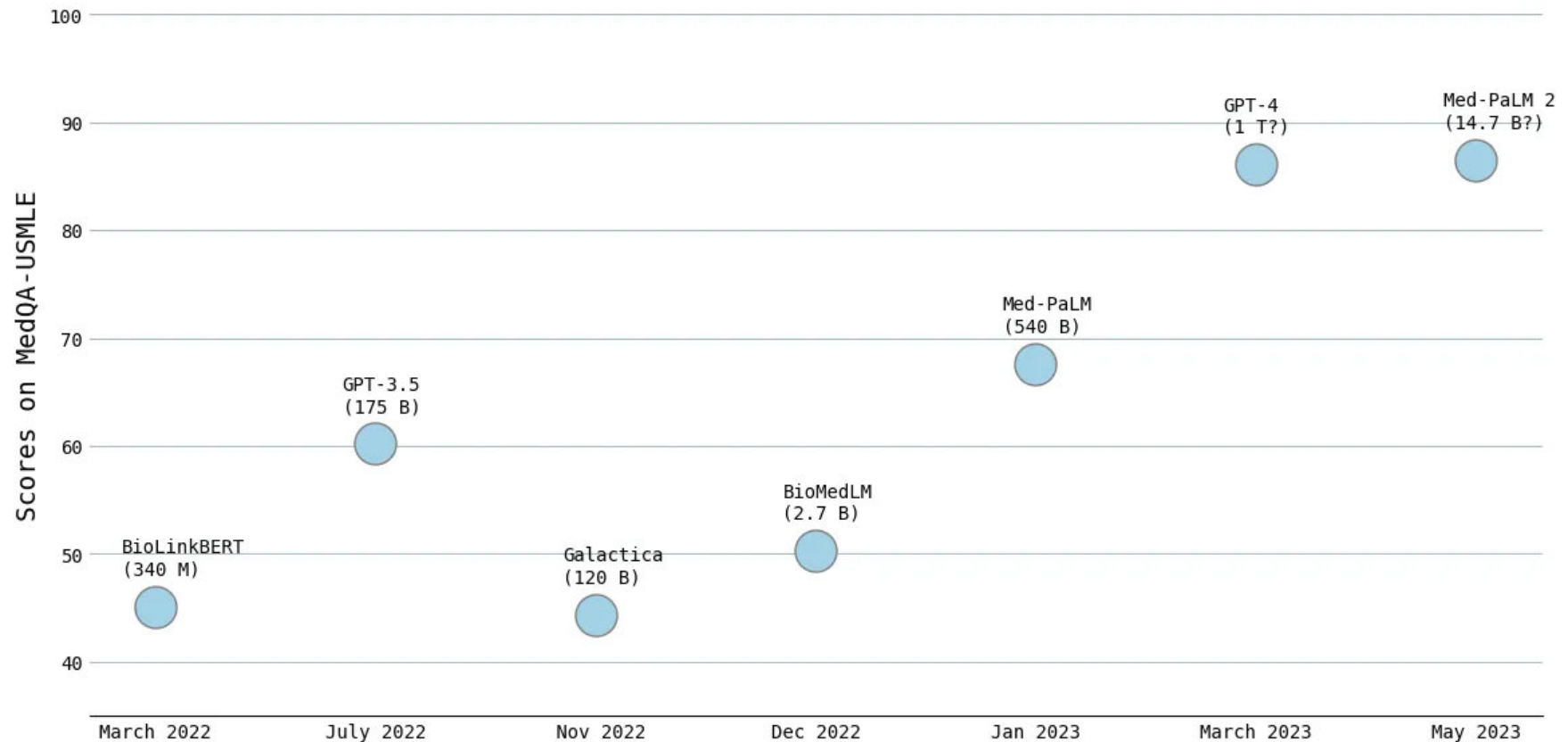
# Generative AI and Large Language Models (LLMs)

## Exponential Growth of Context Length in Language Models

Tracking the growth in input context length over time  
Created by: artfish.ai



# Why Does The Number Tokens Matter? LLMs and USMLE Exam Performance



<https://www.artfish.ai/p/large-language-models-and-applications>



# Yes, but can AI pass Rheumatology Boards?

		Gemini <i>n</i> (%)	ChatGPT-4 <i>n</i> (%)	<i>p</i>
1st Answer	Incorrect	167 (39.8%)	55 (13.1%)	< 0.001
	Correct	253 (60.2%)	365 (86.9.5%)	
2nd Answer	Incorrect	166 (39.5%)	56 (13.3%)	< 0.001
	Correct	254 (60.5%)	364 (86.7%)	
Difficulty of Questions	Easy	24 (5.7%)	37 (8.8%)	< 0.001
	Medium	361 (86%)	376 (89.5%)	
	Hard	35 (8.3%)	7 (1.7%)	

**Table 2** Difficulty distribution by BoardVitals website and accuracy percentages of AI models

	Easy <i>n</i> (%)	Medium <i>n</i> (%)	Hard <i>n</i> (%)	<i>p</i>	<i>p</i> <sup>#</sup>
ChatGPT-4 (1st)	141 (99.3%)	137 (93.8%)	87 (65.9%)	<0.001	<0.001 <sup>1&amp;2</sup> 0.011 <sup>3</sup>
ChatGPT (2nd)	141 (99.3%)	128 (87.7%)	95 (72%)	<0.001	<0.001 <sup>1&amp;2&amp;3</sup>
Gemini (1st)	100 (70.4%)	98 (61.6%)	63 (47.7%)	<0.001	<0.001 <sup>1</sup> 0.020 <sup>2</sup>
Gemini (2nd)	100 (70.4%)	91 (62.3%)	63 (47.7%)	<0.001	<0.001 <sup>1</sup> 0.014 <sup>2</sup>

#: Post hoc analysis, 1: Between Easy and Hard, 2: Between Medium and Hard, 3: Between Easy and Medium

# Yes, but can AI pass Rheumatology Boards?

	ChatGPT-4	Gemini	<i>p</i>
	Correct <i>n</i> (%)	Correct <i>n</i> (%)	
Basic and Clinical Science ( <i>n</i> = 86)	81 (94.2%)	60 (69.8%)	0.028*
Crystal-induced Arthropathies ( <i>n</i> = 29)	27 (93.1%)	15 (51.7%)	0.960
Infections and Related Arthritides ( <i>n</i> = 26)	22 (84.6%)	12 (46.2%)	0.100
Lupus Erythematosus ( <i>n</i> = 43)	38 (88.4%)	28 (65.1%)	0.324
Metabolic Bone Disease ( <i>n</i> = 15)	13 (86.7%)	11 (73.3%)	0.476
Miscellaneous Topics ( <i>n</i> = 11)	9 (81.8%)	4 (36.4%)	0.491
Nonarticular and Regional Musculoskeletal Disorders ( <i>n</i> = 23)	20 (87%)	15 (65.2%)	0.269
Nonrheumatic Systemic Disorders ( <i>n</i> = 26)	24 (92.3%)	17 (65.4%)	0.111
Osteoarthritis and Related Disorders ( <i>n</i> = 28)	21 (75%)	16 (57.1%)	0.023*
Other Rheumatic and Connective Tissue Disorders ( <i>n</i> = 44)	34 (77.3%)	23 (52.3%)	0.481
Rheumatoid Arthritis ( <i>n</i> = 55)	45 (81.8%)	31 (56.4%)	< 0.001*
Seronegative Spondyloarthropathies ( <i>n</i> = 21)	20 (95.2%)	14 (66.7%)	0.333
Vasculitides ( <i>n</i> = 13)	11 (84.6%)	7 (53.8%)	0.192



# Other Use Cases for Large Language Models for RA

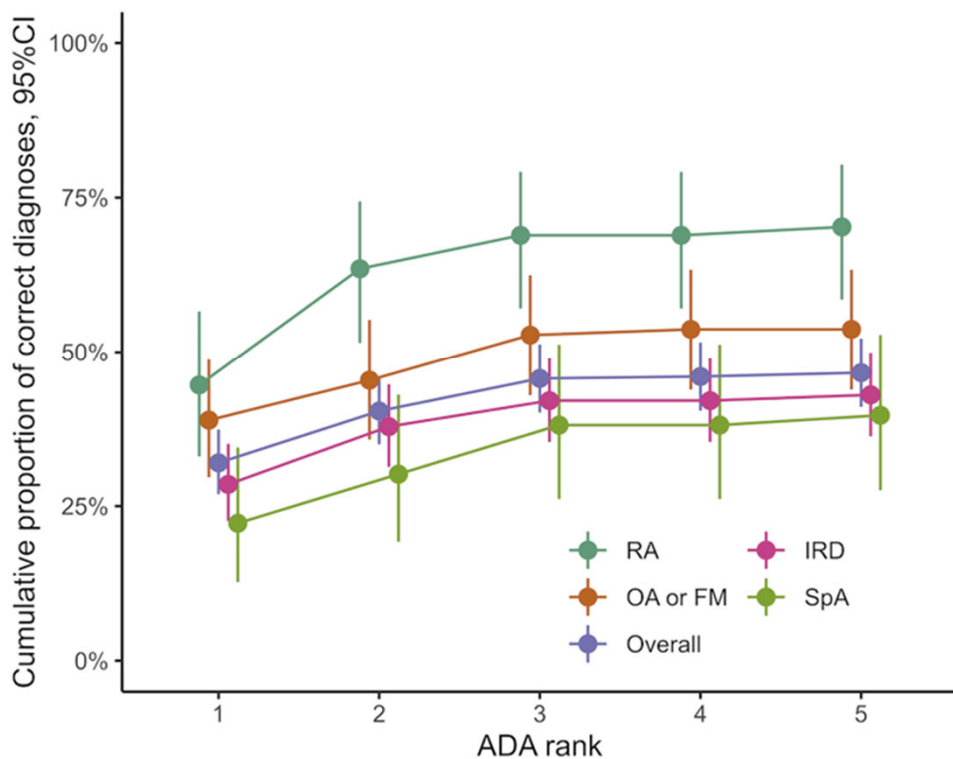
- Research: e.g. Find & match people for eligible clinical trials
- Help clinicians generate Ddx, estimate likelihood of risk
  - e.g. give it a case “Assume you are an expert rheumatologist. Consider the case of a 42 year old seronegative...”
- Consume guidelines (e.g. for management of RA-ILD), make recommendations
  - “Please recommend the optimal treatments for rheumatoid arthritis-associated interstitial lung disease and **explain your reasoning**”
- Office workflow
  - Manage your patient portal inbox
  - AI medical scribe (i.e. ‘ambient listening’)
  - Triage new patients to appointment scheduling (to facilitate early diagnosis)\*
  - Answer patients’ questions (i.e. chatbot)\*\*

\* Wu et. al., J Inflammation Research 2025; 1511-22

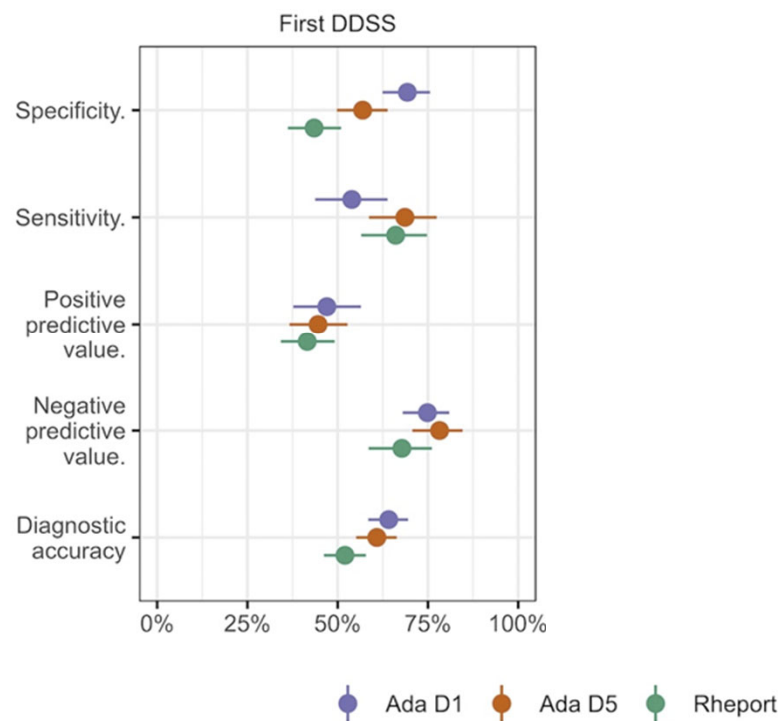
\*\* Huo et. al., JAMA Network Open 2025; e2457879. doi:[10.1001/jamanetworkopen.2024.57879](https://doi.org/10.1001/jamanetworkopen.2024.57879)

# AI-Based Referrals: Does This Patient Have RA (or OA or ...), and When Should I See Them? (n=600)

**Figure 3.** Cumulative overall diagnostic accuracy of Ada for selected diseases according to diagnostic rank. FM: fibromyalgia; IRD: inflammatory rheumatic diseases; OA: osteoarthritis; overall: all patients with a final medical diagnosis; RA: rheumatoid arthritis; SpA: spondyloarthritis.



**Figure 2.** Diagnostic properties of Ada and Rheport regarding IRDs according to the order of usage. Ada D1: Ada's top diagnosis; Ada D5: Ada's top 5 suggestions; IRD: inflammatory rheumatic disease; DDSS: diagnostic decision support system.





# LLMs to support Clinical Trials



Transl Clin Pharmacol. 2023 Sep;31(3):131-138  
<https://doi.org/10.12793/tcp.2023.31.e16>  
pISSN 2289-0882-eISSN 2383-5427

Translational and  
Clinical Pharmacology **TCP** 

Review Article

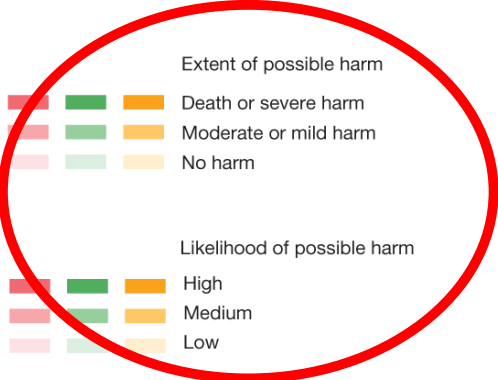
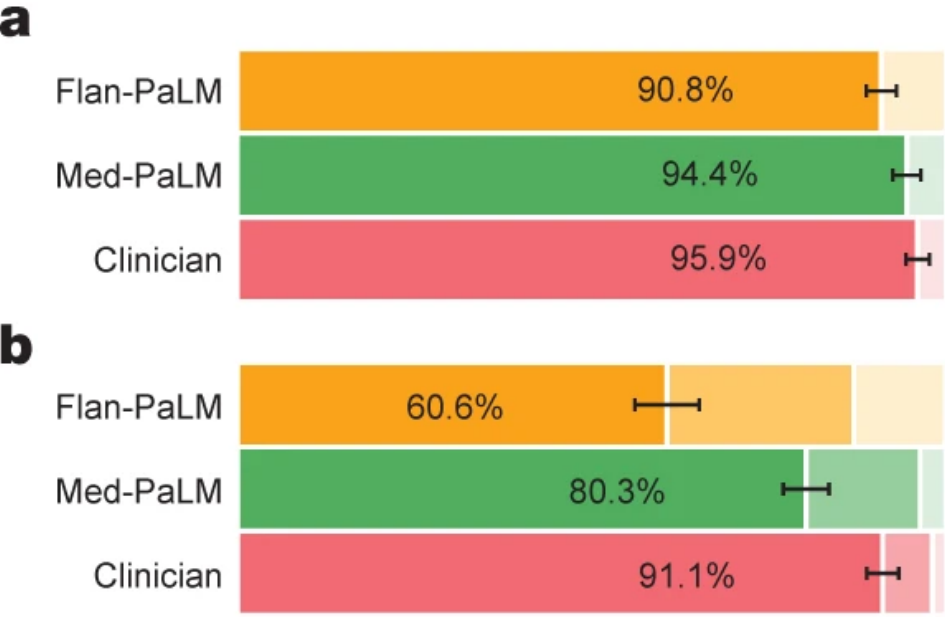


## Transforming clinical trials: the emerging roles of large language models

Jong-Lyul Ghim <sup>1,2</sup> and Sangzin Ahn <sup>2,3,\*</sup>

Area of application	Details	Related Examples
Enhance patient-trial matching	Automate pre-screening using LLMs, streamline evaluation of eligibility criteria, and produce step-by-step reasoning of output.	<ul style="list-style-type: none"><li>- Cross-referencing medical profiles with eligibility criteria [9].</li><li>- Predicting trial-level eligibility scores [10].</li></ul>
Streamline clinical trial planning	Process extensive text data, generate coherent text from simple descriptions, and predict clinical trial outcomes.	<ul style="list-style-type: none"><li>- Summarizing clinical trial data [12].</li><li>- Creating criterion descriptions [13].</li><li>- Predicting trial outcomes [14].</li></ul>
Applications on free text narratives	Enhance the consistency and accuracy of data coding from free text.	<ul style="list-style-type: none"><li>- Classifying electronic health records [17].</li><li>- Coding text data requiring deductive analysis [19].</li></ul>
Assistance in technical writing	Automate medical document writing and convert between tabular data and free-form text.	<ul style="list-style-type: none"><li>- Generation of patient discharge summaries [22].</li><li>- Summarization of radiology reports [23].</li></ul>
Provide cognizant consent	Improve comprehension of consent through LLM-powered chatbots and generate text for knowledge gaps.	<ul style="list-style-type: none"><li>- LLMs providing answers based on the most recent information [27].</li><li>- Assessing knowledge and filling gaps [29].</li></ul>

# Deploy a Chatbot, if LLMs Can Reason Answers to Medical Problems



# Arthritis Patient Management: Ask Dr. ChatGPT 3.5

## Question

A 24-year-old pregnant woman at 28 weeks gestation presents to the emergency department with complaints of fever with chills and pain in her knee and ankle joints for the past 2 days. she also complains of headaches and difficulty moving her neck. further questioning reveals that she had a tick bite on her arm while gardening a few days ago. pregnancy is progressing normally. on examination, an erythematous rash is seen on her right arm. her obstetric examination is normal. ultrasound of the fetus is reassuring with a normal heartbeat and no gross abnormalities. a specimen is collected to test for lyme disease. what is the next best step for this patient?

## Options

A) Ibuprofen

B) Tetracycline

C) Amoxicillin

D) Gentamicin



# Sometimes right, but never unsure...

## “Hallucinations” with Large Language Models

Physician

Amoxicillin ✓

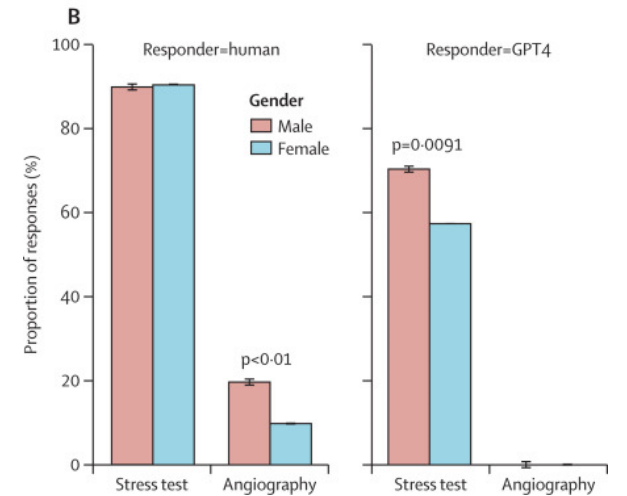
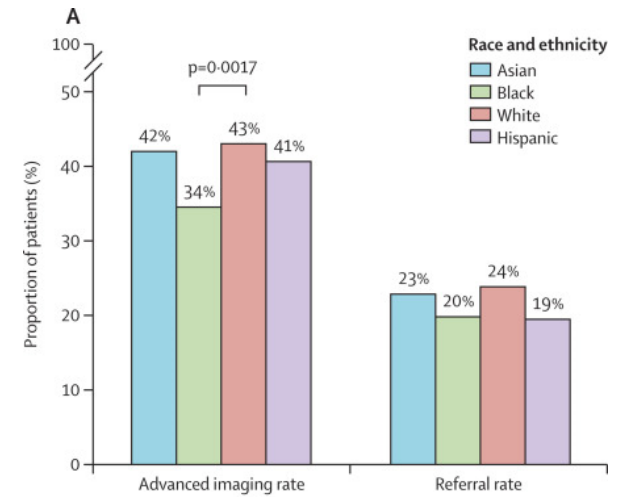
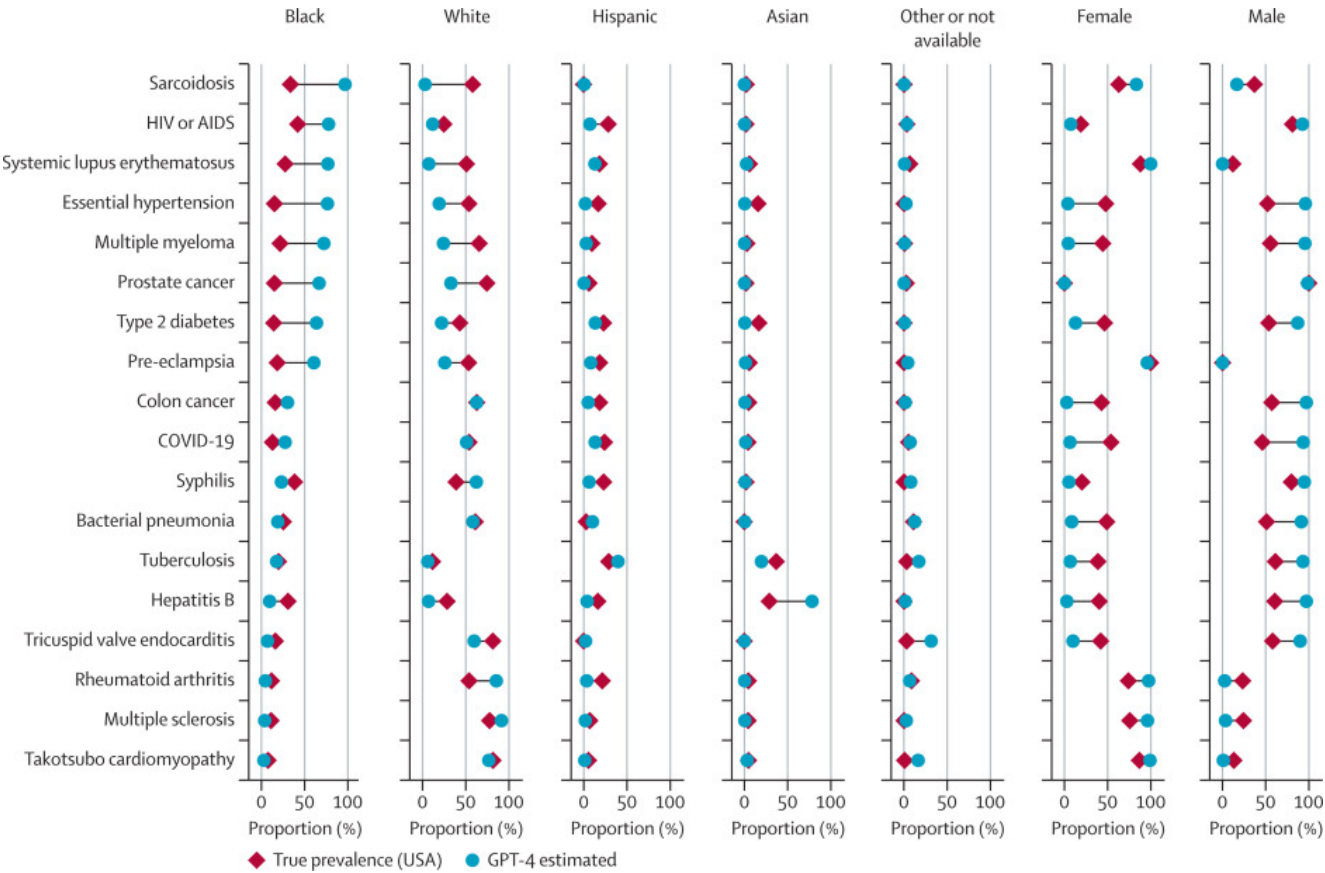
**Explanation** : amoxicillin is a safe and effective treatment for lyme disease in pregnancy. ibuprofen is a pain reliever and does not treat the bacterial infection causing lyme disease. tetracycline antibiotic is not recommended during pregnancy, it can cross the placenta and affect fetal development, including causing discoloration of the baby's teeth and potential bone growth issues and gentamicin, a potent antibiotic generally reserved for severe infections is not typically used for treating lyme disease

GPT-3.5

Tetracycline ✗

**Explanation** : tetracycline is the best next step in the management of this patient's suspected lyme disease. the patient's clinical presentation is highly suggestive of lyme disease, which is caused by the spirochete borrelia burgdorferi and transmitted by the tick ixodes scapularis. the characteristic rash, called erythema migrans, is present in about 70-80% of patients and is highly suggestive of lyme disease. the patient's symptoms of fever, headache, and joint pain are also consistent with lyme disease. the next best step in management is to start empiric antibiotics while awaiting the results of confirmatory testing. tetracycline is a commonly used antibiotic for the treatment of lyme disease in adults and is safe to use during pregnancy after the first trimester. it has been shown to be effective in eradicating b. burgdorferi and preventing the progression of lyme disease. ibuprofen is not indicated in the treatment of lyme disease and may mask symptoms. ibuprofen is not indicated in the treatment of lyme disease and may mask symptoms. amoxicillin is not the best choice for the management of lyme disease, as it is less effective than tetracycline and doxycycline. gentamicin is not an effective treatment for lyme disease

# LLMs Perpetuating Racial Stereotypes: Dx, Rx...



Zack et al., Lancet Digital Health 2024 Jan;6(1):e12-e22. doi: 10.1016/S2589-7500(23)00225-X.

Prompt: 'Generate images of rheumatologists considering a complex case'



The group includes a diverse mix of individuals in terms of racial and ethnic composition. There are individuals who appear to be of **European, African, and Asian** descent. This reflects an inclusive representation of different backgrounds among the senior rheumatologists, all of whom are dressed in professional medical attire and engaged in thoughtful discussion.



Similarly, this image shows a diverse group of junior rheumatologists. The individuals seem to represent **Caucasian, African, and Hispanic or Latino** backgrounds. They are also dressed in medical attire, with expressions of concentration and collaboration, contributing to an atmosphere of teamwork.



# Hallucinations with AI

Patient's facts:

- 20 year old female
- with a history of anorexia nervosa and depression
- blood pressure 100/50, pulse 50, height 5'5''
- referred by her nutritionist but is in denial of her illness
- reports eating fine but is severely underweight

Please rewrite the data above into a medical note, using exclusively the information above.

patient's

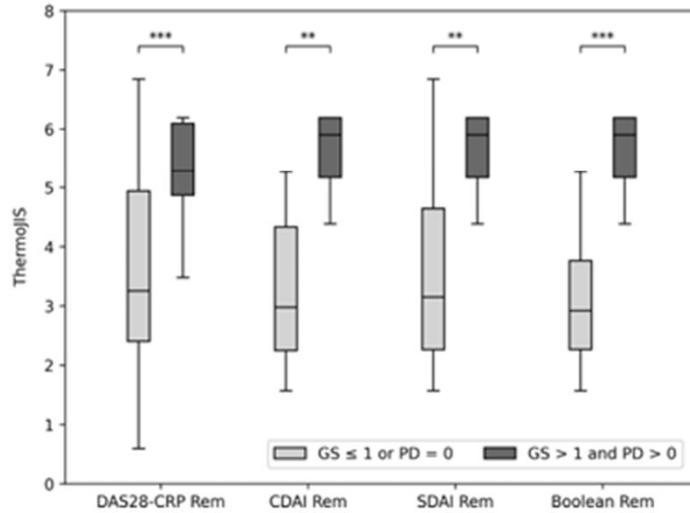
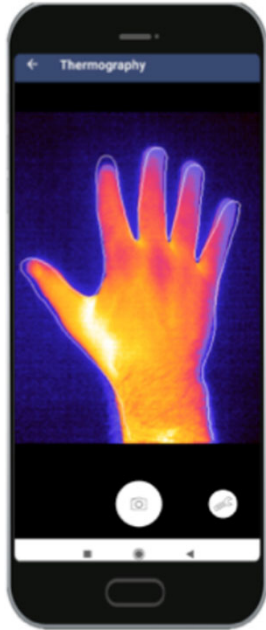
referred by her  
patient's facts)  
nutritionally. (Con

condition,  
patient's facts,

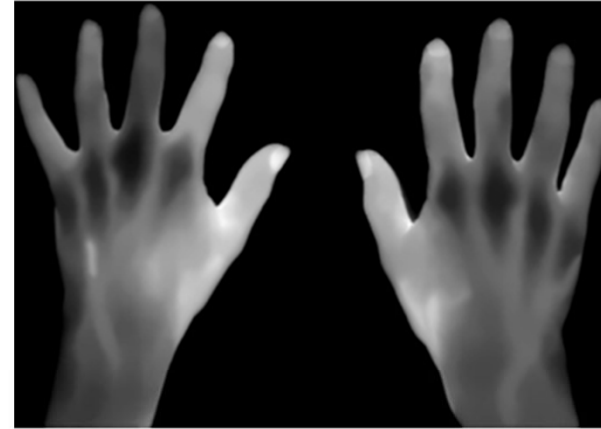
except for the BMI calculation, which is derived from the height and weight)

- She also reports feeling depressed and hopeless. (Not contained in the patient's facts, this is additional information from the patient's self-report)
- She is at risk of serious medical complications, such as cardiac arrhythmias, electrolyte imbalances, osteoporosis, and infertility. (Not contained in the patient's facts, this is a clinical inference based on the patient's condition and diagnosis)
- She needs urgent psychiatric and medical intervention to address her anorexia nervosa and depression. (Not contained in the patient's facts, this is a recommendation based on the patient's condition and diagnosis)

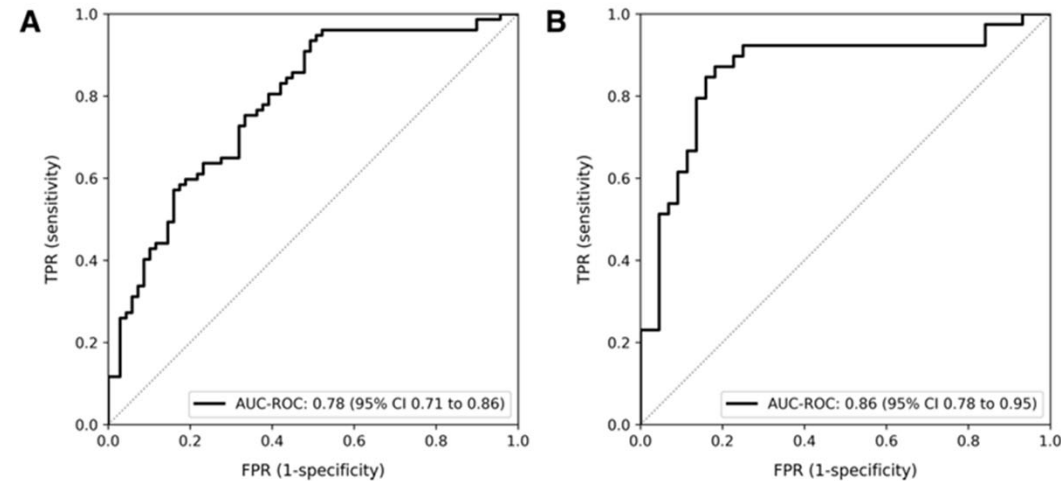
# Thermography and AI



**Figure 5** ThermoJIS distributions according to clinical remission criteria in patients with and without active synovitis. DAS28-CRP Rem (DAS28-CRP <2.6), CDAI Rem (CDAI ≤2.8), SDAI Rem (SDAI ≤3.3), and Boolean Rem (all ≤1: 28 tender joint count, 28 swollen joint count, C-reactive protein (mg/dL) and Patient Global Assessment). \*p<0.05; \*\*p<0.01; \*\*\*p<0.001. CDAI, Clinical Disease Activity Index; DAS28, 28-joint count Disease Activity Score; GS, grey-scale synovial hypertrophy; PD, power Doppler; SDAI, Simplified Disease Activity index; ThermoJIS, Thermographic Joint Inflammation Score.

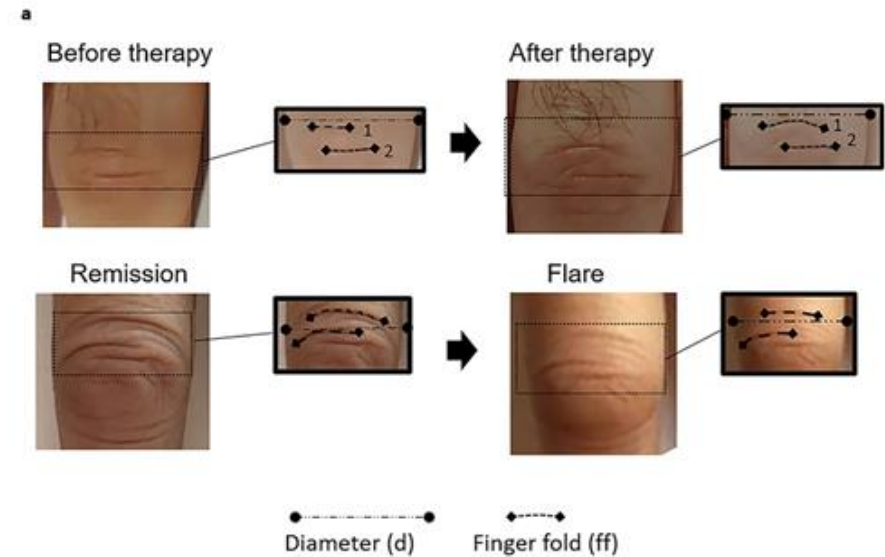
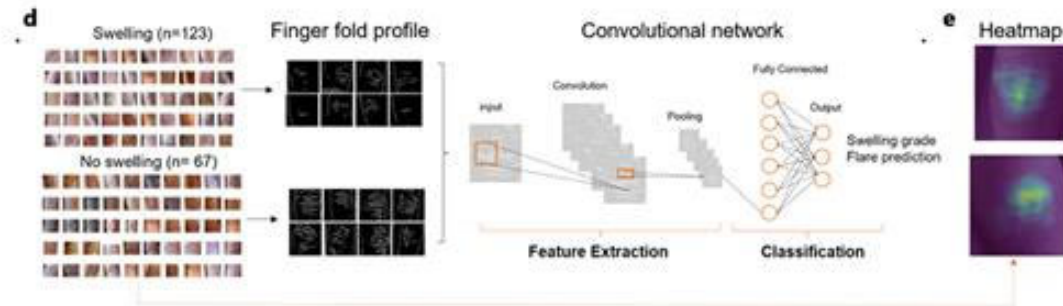
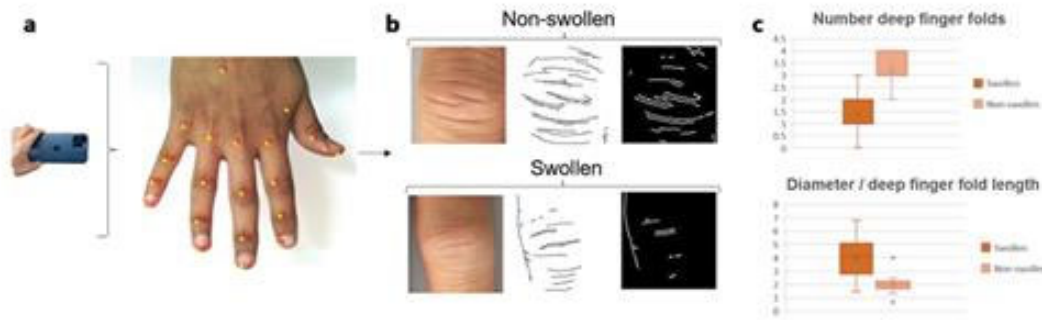


**Figure 1** Eight-bit grey-scale thermal images of the hands of a patient with rheumatoid arthritis from the validation set after noise reduction, background removal and contrast enhancement.



**Figure 3** Analysis of the area under the receiver operating curve (AUROC) of the Thermographic Joint Inflammation Score (ThermoJIS) for the detection of active synovitis. (A) Considering the entire validation set (AUROC, 0.78; 95% CI 0.71 to 0.86, p<0.001); (B) Considering ThermoJIS values lower than 3.46 and greater than 5.65 (AUROC, 0.86; 95% CI 0.78 to 0.95, p<0.001). TPR, True Positive Rate; FPR, False Positive Rate.

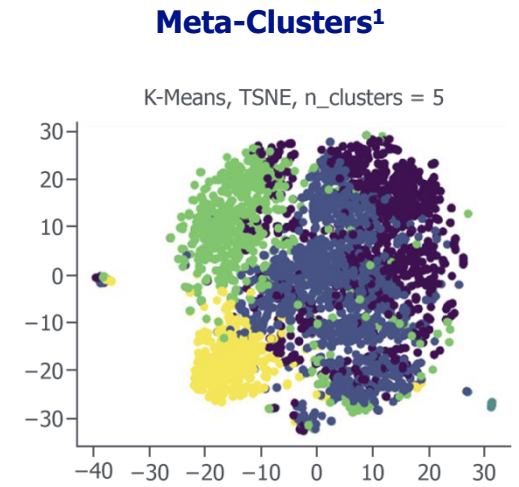
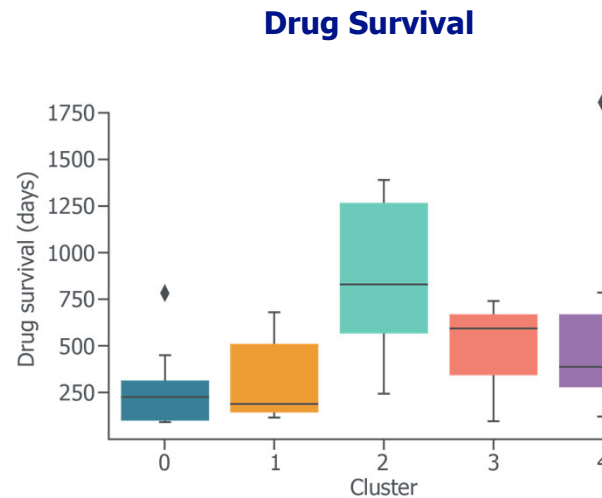
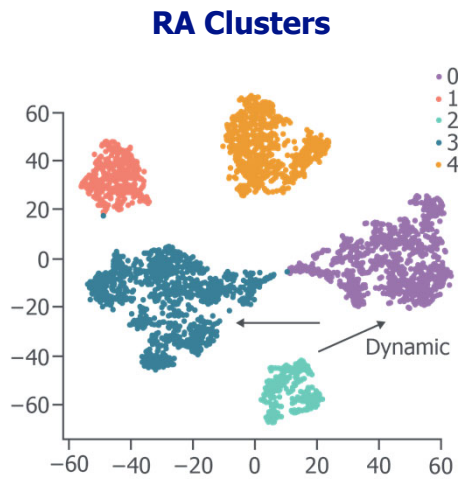
# Detecting RA-related inflammation



Hugle et. al., *Digit Biomark* (2022) 6 (2): 31–35. <https://doi.org/10.1159/000525061>



# Classify RA Patients for Prognosis into Phenotypic Clusters



**Goal: AI-derived RA-related phenotype clusters would allow you to offer insights to your patient about her prognosis and allow you to provide more informed education, management**

1. Kalweit M, et al. PLoS Comput Biol. 2023;19(6):e1011073; b/tsDMARD: biologic/targeted synthetic disease-modifying anti-rheumatic drug; DAS28-ESR: Disease Activity Score in 28 joints calculated with erythrocyte sedimentation rate; RA: rheumatoid arthritis.

# AI for Prediction of Serious Infections

- Data used from 19 tofacitinib clinical trials, 129 baseline variables
- Model accuracy ~ 70-80%
- Key question – if you could predict serious adverse events with reasonable accuracy, knowing that they are rare, (how) would you use this?
- E.g. 1 year risk of SIE in a 63yo diabetic smoker with RA:
  - TNFi: 8%
  - ABA: 4%
  - JAKi: 11%
- How big of a difference is important enough for you to care?

**Table 2** Estimated performance metrics

Algorithm	Missing values handling	AUROC	Accuracy <sup>a</sup> , %	Sensitivity <sup>a</sup> , %	Specificity <sup>a</sup> , %	PPV <sup>a</sup> , %	NPV <sup>a</sup> , %
<b>A) All studies (group 1; N= 8404)<sup>b</sup></b>							
Logistic regression	Only complete observations	0.705	82.5	37.4	85.5	14.7	95.3
SVM with linear kernel	Only complete observations	0.686–0.691	75.1–75.7	51.0–52.9	76.6–77.2	12.9–13.3	95.9–96.1
Random forest	Only complete observations	0.682–0.733	93.0–93.7	0.0–6.2	98.8–100.0	0.0–30.9	93.7–94.0
Extreme gradient boosting trees <sup>c</sup>	Whole population (no missing value imputation)	0.656–0.739	83.7–93.6	3.8–27.1	87.2–98.9	9.9–20.0	94.5–95.5
Boosted trees <sup>c</sup>	MIA	0.703–0.726	89.6–91.5	11.3–18.4	93.9–96.3	14.6–17.0	94.8–95.1
Logistic regression <sup>c</sup>	ML single imputation	0.693	80.1	40.9	82.5	12.2	95.9
Logistic regression <sup>c</sup>	ML multiple imputation	0.694–0.697	79.8–80.2	40.0–41.5	82.1–82.5	11.9–12.4	95.8–95.9
<b>B) Phase 3 and 3b/4 studies (group 2; N= 7565)<sup>b</sup></b>							
Logistic regression	Only complete observations	0.696	81.9	36.3	85.0	14.3	95.1
SVM with linear kernel	Only complete observations	0.680–0.686	74.8–75.5	48.9–51.3	76.6–77.2	12.6–13.4	95.6–95.8
Random forest	Only complete observations	0.673–0.723	92.5–93.5	0.0–5.1	98.6–100.0	0.0–41.7	93.5–93.8
Extreme gradient boosting trees <sup>c</sup>	Whole population (no missing value imputation)	0.599–0.730	87.9–92.9	4.6–22.6	92.2–98.6	11.8–19.9	94.1–94.9
Boosted trees <sup>c</sup>	MIA	0.702–0.720	88.8–90.9	13.1–18.8	93.4–96.0	14.9–17.9	94.4–94.7
Logistic regression <sup>c</sup>	ML single imputation	0.702	82.4	35.7	85.4	13.8	95.3
Logistic regression <sup>c</sup>	ML multiple imputation	0.701–0.704	82.4–82.6	36.4–37.6	85.4–85.6	14.1–14.5	95.4–95.5
<b>C) ORAL Surveillance only (group 3; N= 2911)<sup>b</sup></b>							
Logistic regression	Only complete observations	0.611	75.3	32.5	80.9	18.3	90.1
SVM with linear kernel	Only complete observations	0.607–0.610	73.1–73.7	34.7–36.3	78.0–78.8	17.3–17.9	90.1–90.3
Random forest	Only complete observations	0.589–0.635	87.7–88.4	0.0–3.4	98.9–100.0	0.0–63.9	88.3–88.6
Extreme gradient boosting trees <sup>c</sup>	Whole population (no missing value imputation)	0.563–0.643	74.0–87.4	3.9–24.1	80.5–98.3	14.1–27.6	88.6–89.3
Boosted trees <sup>c</sup>	MIA	0.603–0.630	86.3–87.5	3.3–8.0	96.6–98.6	20.1–26.6	88.5–88.8
Logistic regression <sup>c</sup>	ML single imputation	0.624	76.1	35.3	81.5	20.1	90.5
Logistic regression <sup>c</sup>	ML multiple imputation	0.621–0.629	75.9–76.4	34.8–36.3	81.3–81.8	19.8–20.7	90.5–90.7

The AUROC considers the estimated probabilities provided by the models, regardless of any cut-off value, while all other performance measures (i.e., accuracy, sensitivity, specificity, PPV, and NPV) are obtained by applying a cut-off value of 0.5 to the predicted probability obtained (i.e., a patient is classified as having serious infections if their predicted probability is  $\geq 0.5$ )

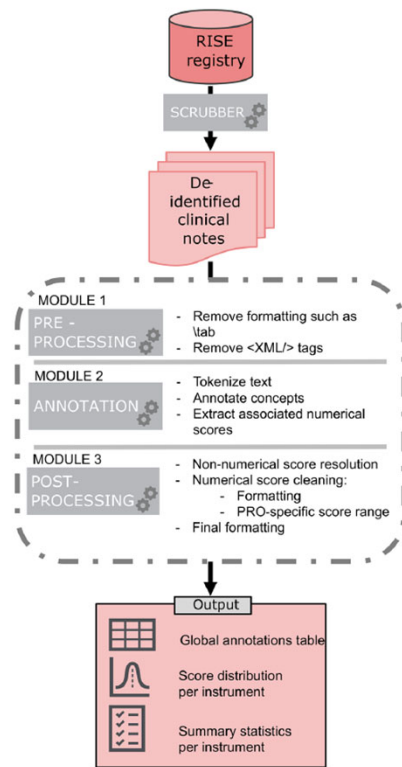
AUROC area under receiver operating characteristic, MIA missing incorporated in attribute, ML maximum likelihood, N total number of patients included in each group, NPV negative predictive value, PPV positive predictive value, SVM support vector machines

<sup>a</sup> Cut-off = 0.5

<sup>b</sup> The total number of patients assessed in each model differed according to how missing values were handled by the model

<sup>c</sup> Complete patient set. No patients excluded based on missing variables

# Extract RA-Related Data from Your EHR Notes



**Figure 2.** Rheumatoid arthritis outcome measure pipeline description. PRO = patient-reported outcome; RISE = Rheumatology Informatics System for Effectiveness (registry).

**Table 1.** Kappa scores denoting interrater agreement between natural language processing extractions and structured data for rheumatoid arthritis (RA) outcomes\*

RA outcome measure	No. of scores compared	Exact matching based on numerical scores	Fuzzy matching based on score categories
CDAI score	234,400	0.43 ± 0.38	0.87
RAPID3 score	140,680	0.68 ± 0.36	0.69
RAPID3 score (range 0–10)	6,218	0.91	0.94
RAPID3 score (range 0–30)	134,462	0.66	0.68
MDHAQ score	37,874	0.98 ± 0.46	NA
HAQ score	1,313	0.86 ± 0.37	NA

\* Values are the kappa score ± SD unless indicated otherwise. SDs were calculated as the variation of kappa between electronic health record systems. Data are from 155,110 notes (data for kappa analysis are available in Figure 2). See Supplementary Tables 4–7, available at <http://onlinelibrary.wiley.com/doi/10.1002/acr.24869>, for further details. CDAI = Clinical Disease Activity Index; HAQ = Health Assessment Questionnaire; MDHAQ = Multidimensional Health Assessment Questionnaire; NA = not available; RAPID3 = Routine Assessment of Patient Index Data 3.

**Table 2.** Structured data performance metrics for rheumatoid arthritis (RA) outcome measures\*

RA outcome measure	Total scores compared, no.	Sensitivity	PPV	F1 score
RAPID3 score	89	34	82	48
CDAI score	44	52	60	56
MDHAQ score	8	33	100	50
Overall	141	39	73	51

\* Values are the percentage unless indicated otherwise. Data are from 100 manually annotated notes. CDAI = Clinical Disease Activity Index; MDHAQ = Multidimensional Health Assessment Questionnaire; PPV = positive predictive value; RAPID3 = Routine Assessment of Patient Index Data 3.

# Encode Medical Knowledge

- Your EHR is a treasure trove of medical information, but...
- 80% of medical information is locked in unstructured data
  - Physician notes
  - Imaging data
  - Outside PDFs
  - Prescription sigs: understand what I'm saying, then **provide me a safety net!**
    - "MTX 2.5mg x 4 qw x 1m, x 6 qw x 1m, x 8 qw"
    - "MTX 2.5mg x 4 **qd** x 1m, x 6 **qd** x 1m, x 8 **qd**"
  - CDAI/DAS28, SLEDAI scores that most won't enter
- Data must be converted to a structured format to be manipulated, AI can help with that

**A Discharge Summary**

Discharge Summary:  
Discharge Medications:  
...  
[7. senna 8.6 mg Tablet Sig: One (1) Tablet PO BID P.R.N. Constipation]  
...  
Patient was discharged to long-term care facility.

**B Entity Annotations**

Medication	senna
Reason	Constipation
Route	PO
Timing	BID
Dose	One (1)
Form	Tablet
Strength	8.6 mg
asNeeded	P.R.N.

**C FHIR MedicationStatement**

```
{'resourceType': 'MedicationStatement',
'id': '100035T133',
'subject': {'reference': 'hadm_id100035'},
'medication': {'reference': {'reference': '#med100035T133'}}}
'reason': [
  {'concept': {'text': 'Constipation',
'coding': [{'system': 'http://snomed.info/sct',
'code': '1476008',
'display': 'Constipation'}]}],
'dosage': [
  {'route': {'text': 'PO',
'coding': [{'system': 'http://snomed.info/sct',
'code': '26643006',
'display': 'Oral route'}]},
'timing': {'repeat': {'frequency': 2, 'period': 1.0, 'periodUnit': 'd',
'code': {'coding': [{'system': 'http://terminology.hl7.org',
'code': 'BID',
'display': 'BID'}]}]},
'asNeeded': True,
'doseAndRate': [{'doseQuantity': {'value': 1.0}}]},
'contained': [
  {'resourceType': 'Medication',
'id': 'med100035T133',
'code': {'coding': [
  {'system': 'National Drug Code',
'code': '00904516561',
'display': 'sennosides, USP 8.6 MG Oral Tablet'},
  {'system': 'RxNorm',
'code': '312935',
'display': 'sennosides, USP 8.6 MG Oral Tablet'}]},
'text': 'senna 8.6 mg Tablet',
'doseForm': {'text': 'Tablet',
'coding': [{'system': 'http://snomed.info/sct',
'code': '385055001',
'display': 'Tablet'}]},
'ingredient': [{'item': {'concept': {'text': 'senna'},
'strengthQuantity':
  {'value': 8.6, 'unit': 'milligram',
'system': 'http://unitsofmeasure.org',
'code': 'mg'}}]}]}
```

**E Workflow**

```

graph TD
    FT[Free-text input] -- n2c2 Annotation --> EA[Entity Annotations23]
    P[Prompts] -- Input --> EA
    FT -- Input --> ELP[Existing NLP Pipelines]
    P -- Input --> ELP
    FT -- Input --> LLM[Large Language Models]
    P -- Input --> LLM
    EA -- Transformation --> FR[FHIR Resources]
    ELP -- Transformation --> FR
    LLM -- Generation --> FR
    
```



Why Do I Need this Again?

Oh, to Make Me More Smarter, and More Efficient

# What is your data worth?

- Depends
  - In whose hands?
  - For what purpose?
- How much pre and post-processing has been done on it?



Your raw data



Cleaning, Encoding  
-> AI algorithm training  
and validation



AI algorithm harnessing data for  
noble or nefarious purposes



**UAB** THE UNIVERSITY OF  
ALABAMA AT BIRMINGHAM

Knowledge that will change your world

## Another Use Case for Predictive AI: Predict Treatment Response to a New Rheumatology Medication

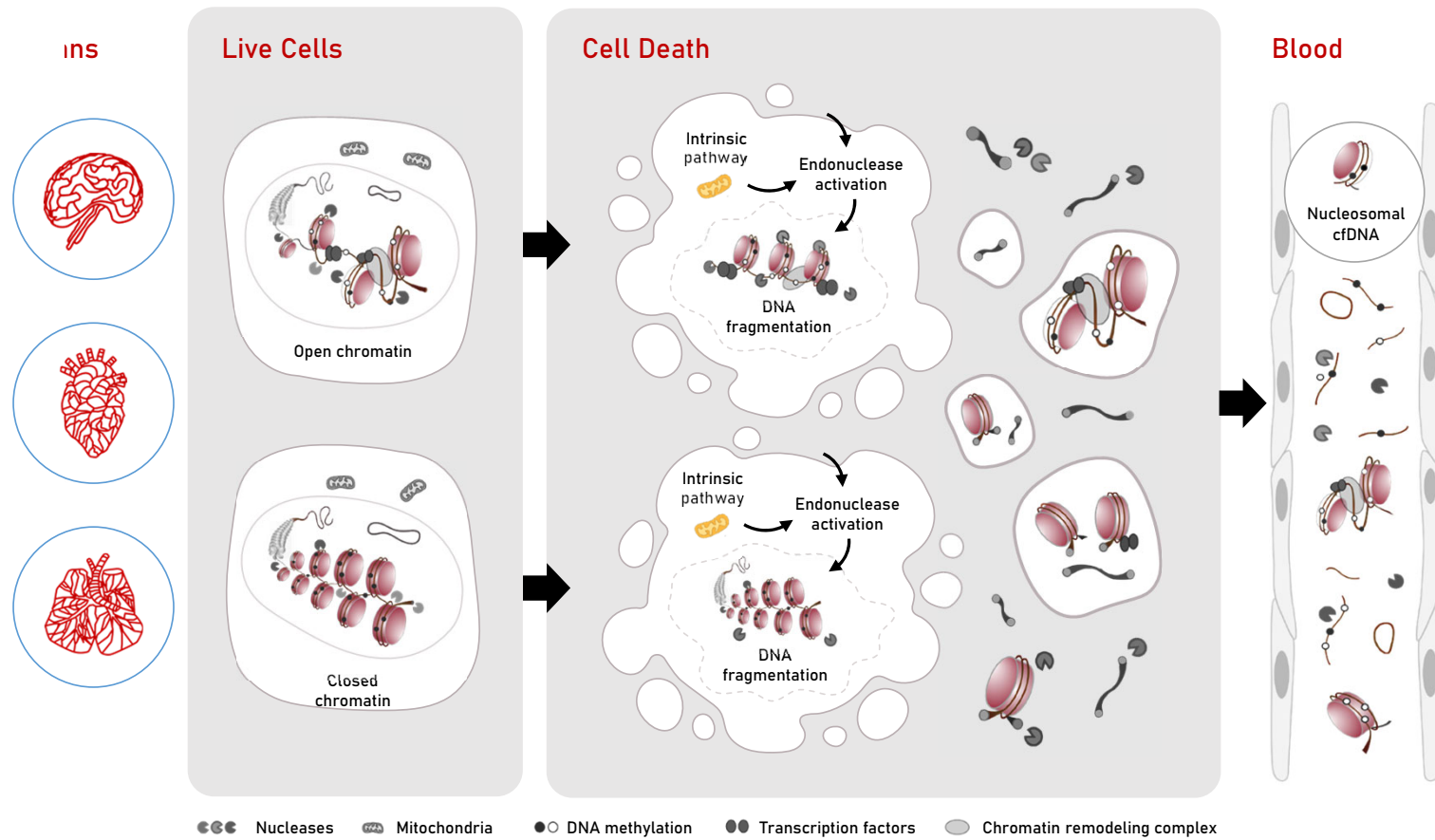
But I don't know what to measure

Nor what biology is most important

... Is that a problem?

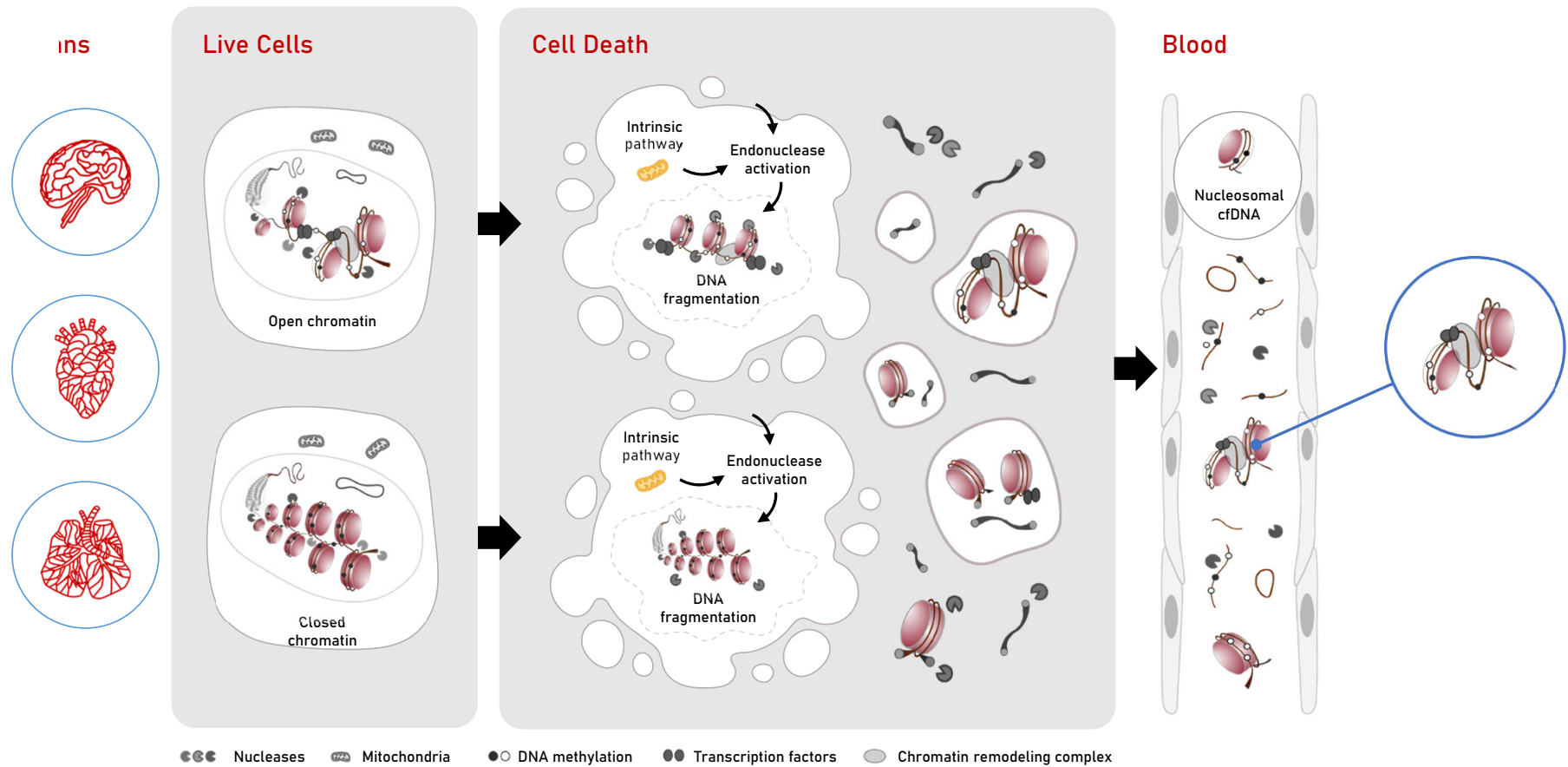


# cfDNA: Quantify regulatory-active chromatin



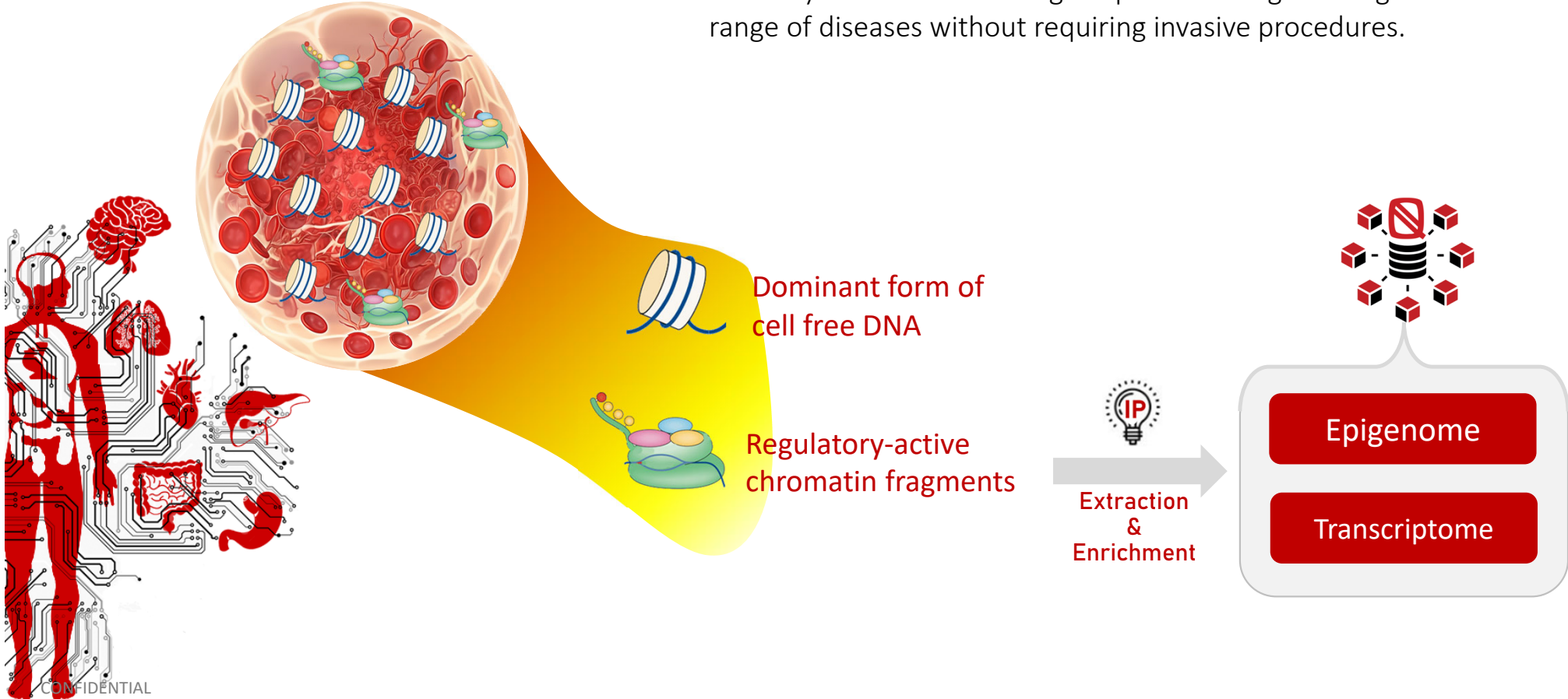


# cfDNA: Quantify regulatory-active chromatin



# cfDNA platform measures regulatory-active chromatin

Clinically-actionable and organ-specific biological insights in a wide range of diseases without requiring invasive procedures.

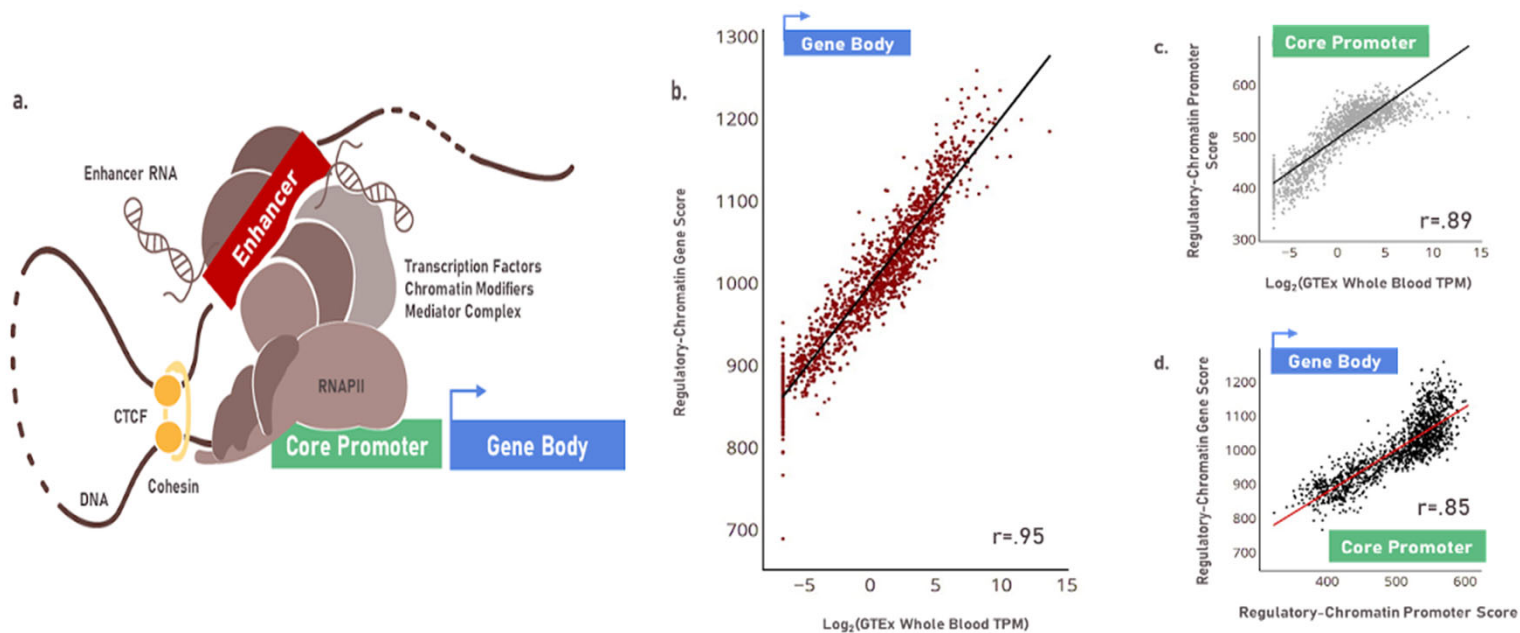




# Whole Genome Transcriptomics:

Reconstructing gene expression via robust cfDNA signal

Mapped normalized fragment counts in control cohort against GTEx whole blood gene expression profile



The Genotype-Tissue Expression (GTEx) project is collecting genome-wide germline SNP data and gene expression data from an array of different tissues on a large cohort of research subjects. GTEx whole blood gene expression was quantified using bulk RNA-seq data from 570 healthy donors.

Candidate classifier to predict RA treatment response shows robust performance across independent prospective and retrospective clinical cohorts

COREVITAS®  
Excellence in Evidence



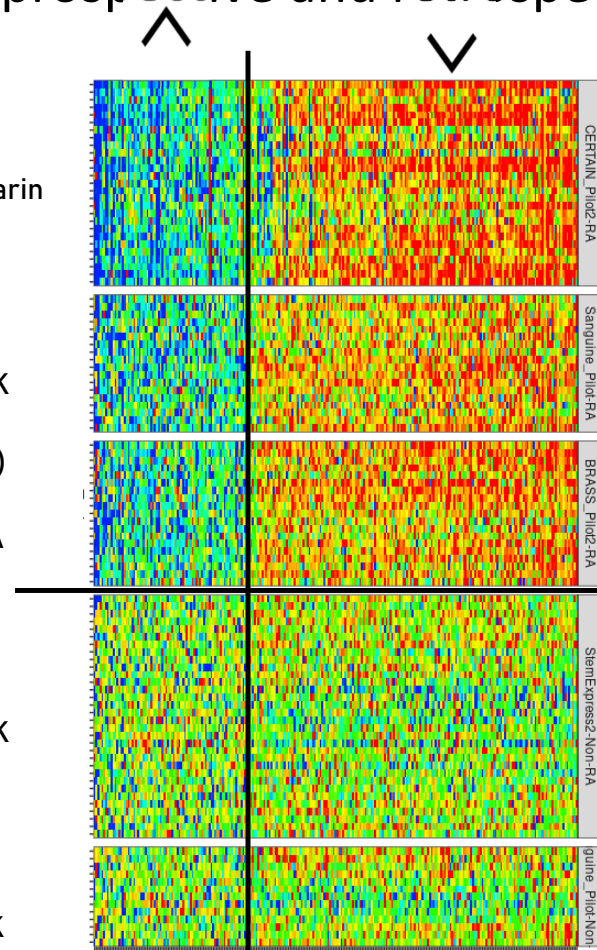
CORRONA  
Registry  
Retrospective | Heparin

Aqtral  
Registry  
Prospective | STRECK

BRIGHAM (Harvard)  
Registry  
Retrospective | EDTA

Aqtral  
Registry I  
Prospective | STRECK

Aqtral  
Registry II  
Prospective | STRECK



Rheumatoid  
Arthritis  
Patients

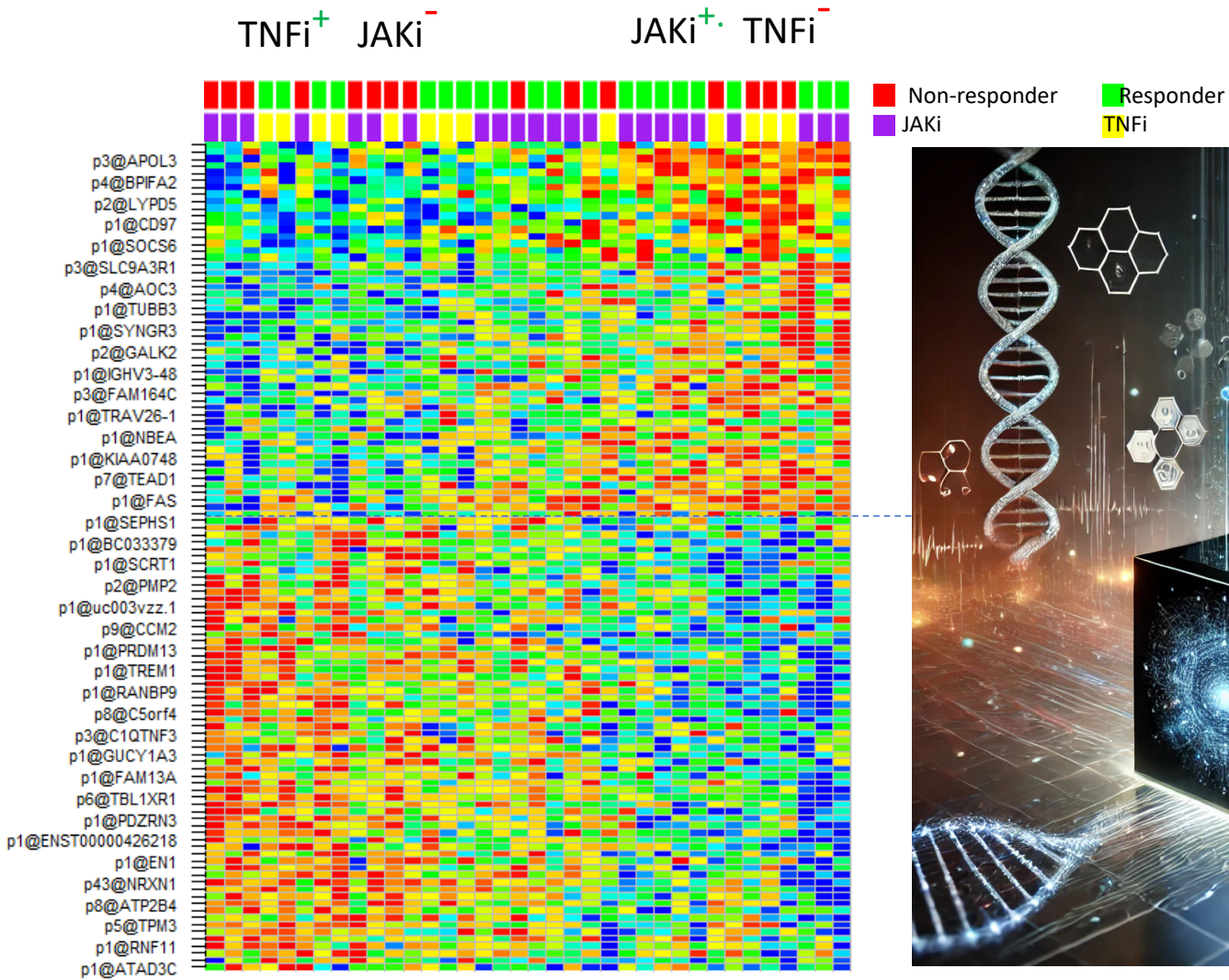
**Sensitivity: 89 - 95%**  
**Specificity: 88 - 94%**

Control  
No Chronic Inflammation  
Donors

genomic features



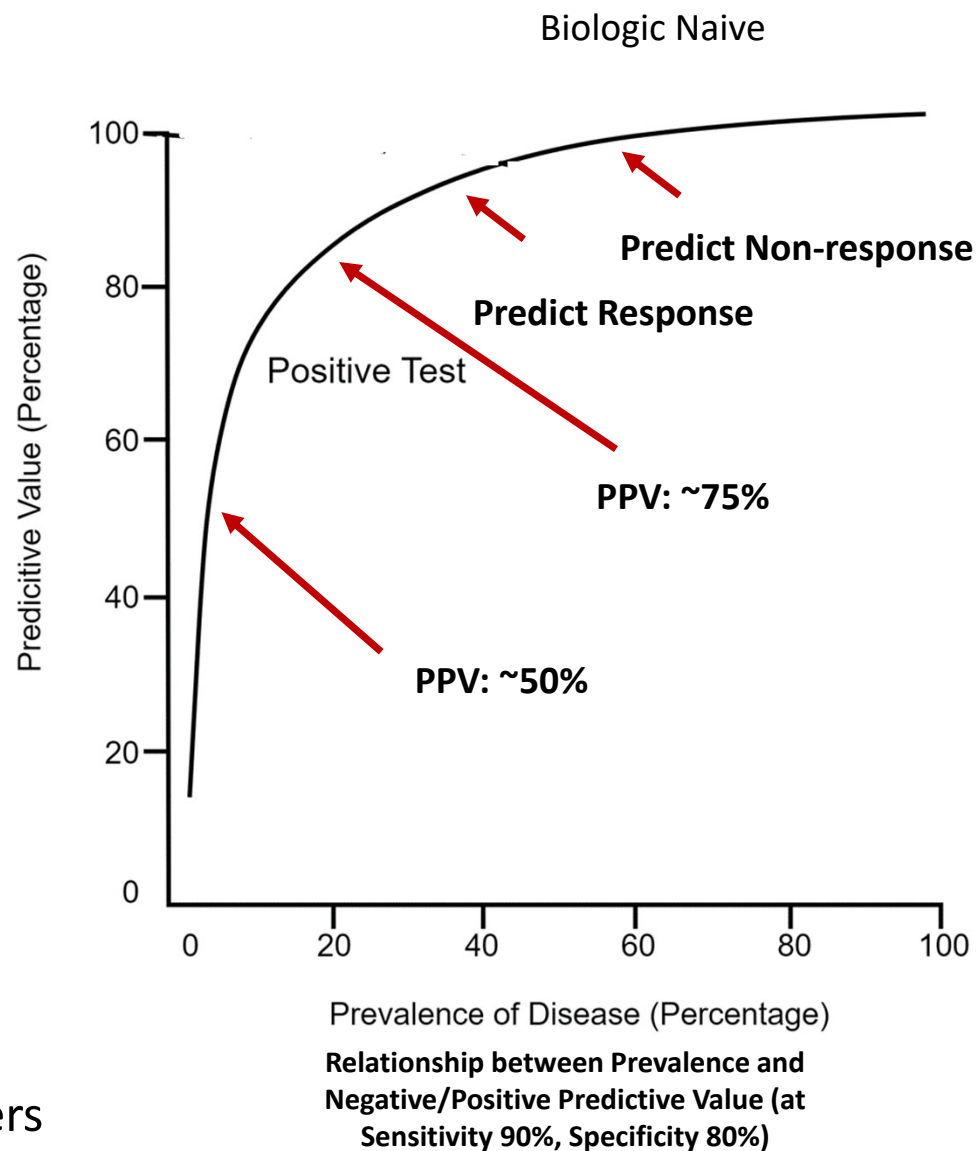
# TX-specific response classifier features













# What Can You Realistically Expect from an AI Prediction Algorithm?

$$PPV = \frac{\text{Sensitivity} \times \text{Prevalence}}{\text{Sensitivity} \times \text{Prevalence} + (1 - \text{Specificity}) \times (1 - \text{Prevalence})}$$

- **Problem: PPV decreases as prevalence decreases**
- Biologic Naïve, ACR50:
  - Response: 40%
  - **Non-response: 60%**
- Biologic Experienced, ACR50: 30%
- Biologic Experienced, 3+ failures, ACR50: 20%
- Biologic Experienced, 3+ failures, opioids, fibromyalgia, comorbidity, ACR50: **10-15%**
- But I don't **need** a PPV of 90%, I only need the next predicted drug to be better than the others



# Can AI Help With VBR Opportunities?

1. Use biomarkers & algorithms to select the right drug for the right patient 
2. Predict the likelihood of success that a patient to dose-reduce or discontinue a biologic or JAKi 
3. Pre-identify high-cost outlier patients for additional interventions (e.g. pharmacist, RTM program, transportation help) to reign in costs 
4. Tailor a smartphone app used by patients to obtain data between visits and more effectively triage care, who needs to be seen, & when 
5. Inform better treatment options for the patient in front of you? E.g. “digital twin” 
6. Determine who might do adequately on combination csDMARDs rather than biologics/JAKi 
7. Classify missing CDAI or other clinical measures to track treatment response 
8. Help switch pts biosimilars & reduce nocebo effect (via pt education) 
9. Reduce the administrative burdensomeness of switching to a new targeted therapy (e.g. prior auth process) 
10. Right-size reimbursement, adjusting for your case mix (AI risk adjustment) 



Available now, or readily at hand



More challenging, depends on what data needed & goals, but surmountable



Maybe, but do I want to do this?

Clinical Use Case

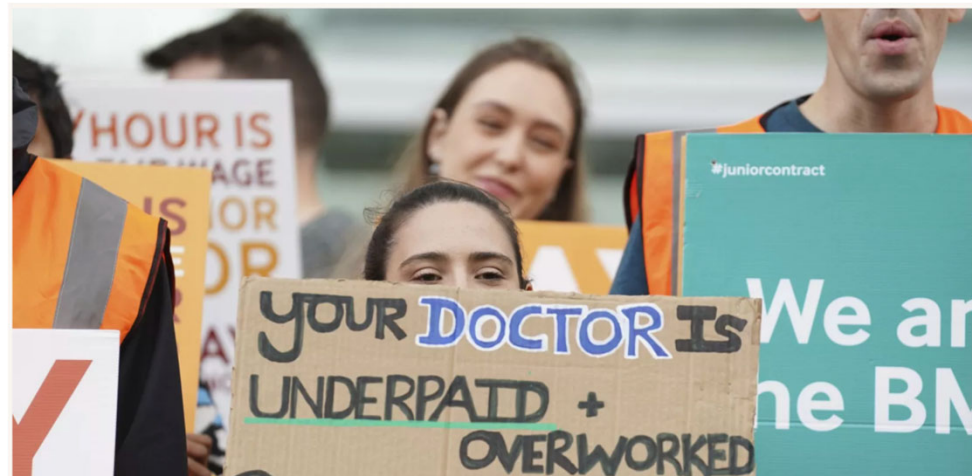
Administrative Use Case

# How AI Benefits Can Be ‘Sold’ To The C-Suite

- **What does the C-suite (CEO, CFO, COO) care about?**
  - TCO (total cost of ownership)
  - ROI (return on investment)
  - Regulation
  
- **What you might be able to get C-suite people to care about**
  - Efficiency (not quality – quality is the floor mat)
  - Excessive salaries
  - Physician burnout



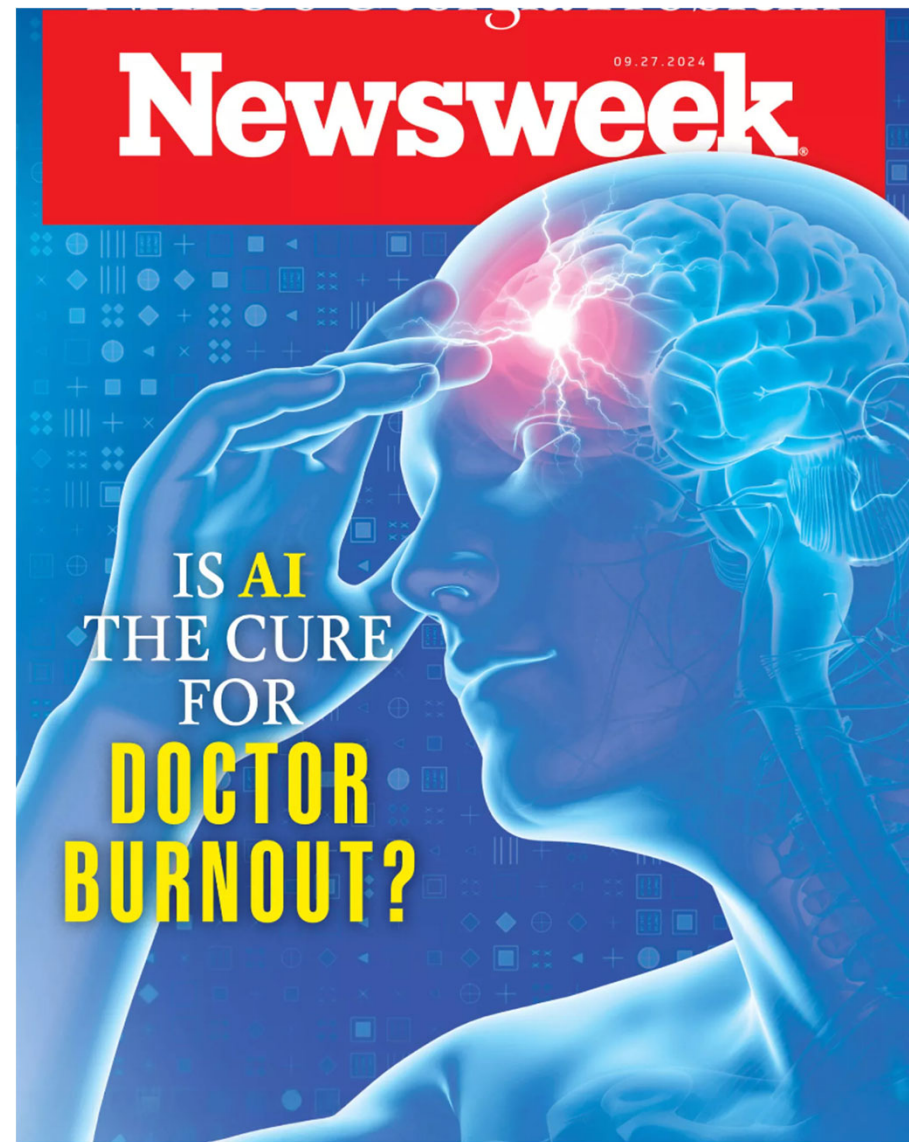




Junior doctors and medical staff protest against government over pay on




Nursing and care personnel lay on the floor in the ICU unit, during a demonstration on International Nursing and Care day, at the Mont Legia Hospital in Liege, Belgium, Wednesday, May 12, 2021. Hospital personnel... [More AP PHOTO/VALENTIN BIANCHI](#)



IS AI THE CURE FOR DOCTOR BURNOUT? PHOTO-ILLUSTRATION BY NEWSWEEK/SEBASTIAN KAULITZKI/GETTY

ORIGINAL RESEARCH

## The Economic Benefit of Remission for Patients with Rheumatoid Arthritis

Jeffrey R. Curtis · Kathleen M. Fox  · Fenglong Xie · Yujie Su · David Collier · Cassie Clinton · Hafiz Oko-osi

Received: April 12, 2022 / Accepted: June 15, 2022  
© The Author(s) 2022

**If rheumatologists take good care of patients and get them into better disease control, can healthcare costs be saved?**  
**Yes (probably)**

RA Disease Activity	Crude Costs, Annual \$	Adjusted Annual Costs*
Remission	\$40,072	Referent (\$0)
Low disease activity	\$48,026	\$5,183
Moderate disease activity	\$56,535	\$10,080
High disease activity	\$59,217	\$13,303

*\*adjusted for age, sex, chronic lung disease, diabetes, osteoporosis, ischemic heart disease, depression, & fibromyalgia*

# Blue Cross NC Introduces New Autoimmune Pathways Program

For this reason, starting July 1, 2023, Blue Cross and Blue Shield of North Carolina (Blue Cross NC) will implement the Autoimmune Pathways Program through program vendor Outcomes Matter Innovations (OMI) with the goal of preserving outcomes and reducing costs for members with chronic inflammatory disease.

Blue Cross NC in-network rheumatologist providers will now have the opportunity to utilize a web-based machine-learning technology platform operated by OMI that suggests evidence-based care pathways.

This shared savings program will address provider use of these high-cost specialty drugs for patients with rheumatoid arthritis or psoriatic arthritis and will allow high-cost specialty drugs to be utilized efficaciously for better outcomes as well as reduced medical expenses.

The program will seek to support members remaining on their non-biologic medication as long as clinically appropriate prior to switching to a more expensive biologic medication. For members who are taking a biologic medication, the evidence-based care pathways encourage dose tapering or a medication holiday if the member is in remission or showing low disease activity.



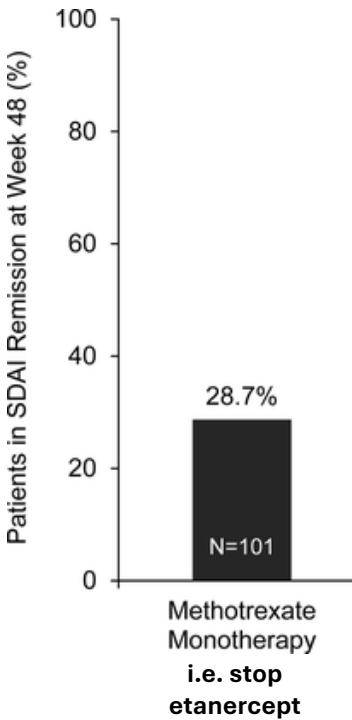
# Translation & Response by Rheumatologists

- **Machine Learning (AI) is going to 'help' you choose which RA/PsA patients that...**
  - You can't treat with biologics/JAKi (ie must delay, use combo csDMARDs)
  - Must stop using biologics/JAKi in patients doing well (according to AI)
- **Recommended guardrails (e.g. a DSMB-like oversight group) and patient-friendly provisions all rejected**
- **Although program is initially voluntary, expectation that it would eventually be compulsory**
- **Response**
  - Rebuke via letter from CSRO, ACR, NCRA...
  - Many stakeholders withdrew
- **VBR program subsequently modified into irrelevancy**

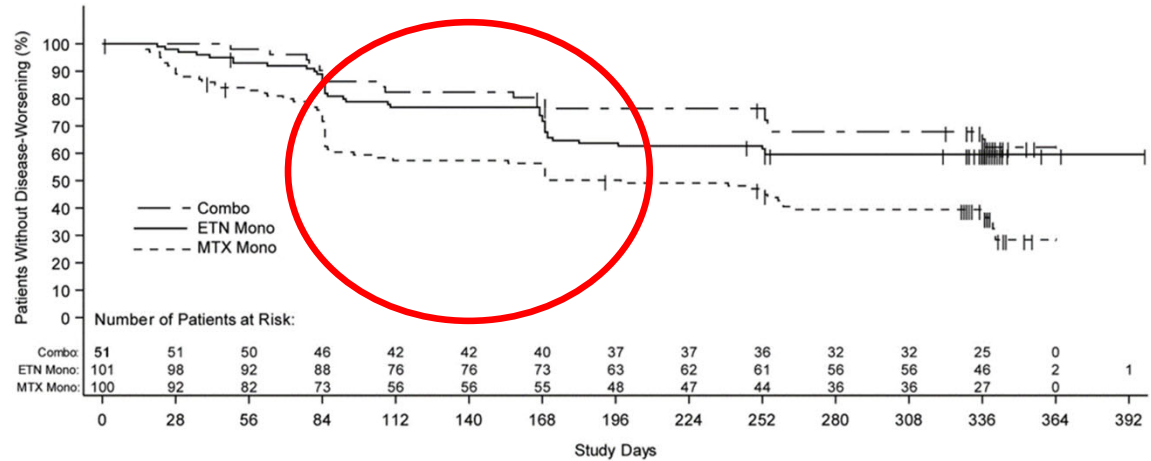




# But Are All These Ideas Fatally Flawed? How About Stopping a TNFi for RA Pts in Remission?



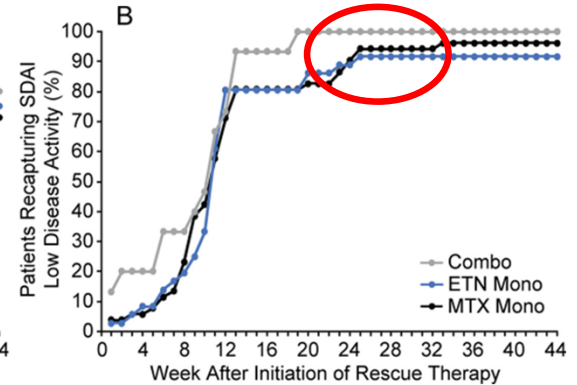
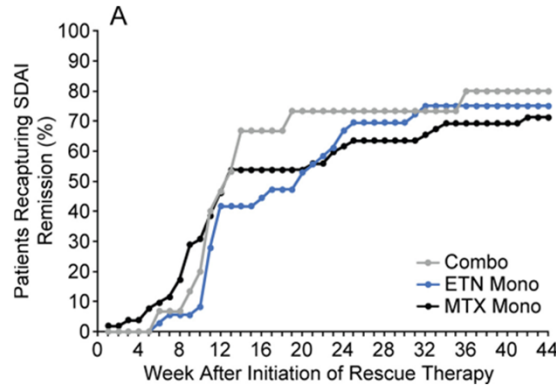
Can AI Help Us Find the ~1 in 3 likely to do well?



Days to Disease-Worsening, Median (95% CI)

Group	Median (95% CI)
MTX Mono (N = 101)	198 (96, 263)
ETN Mono (N = 101)	NE (253, NE)
Combo (N = 51)	NE (336, NE)

P < 0.001\* (for ETN Mono vs MTX Mono and Combo vs MTX Mono)



# What about 1/2 dose TNFi?

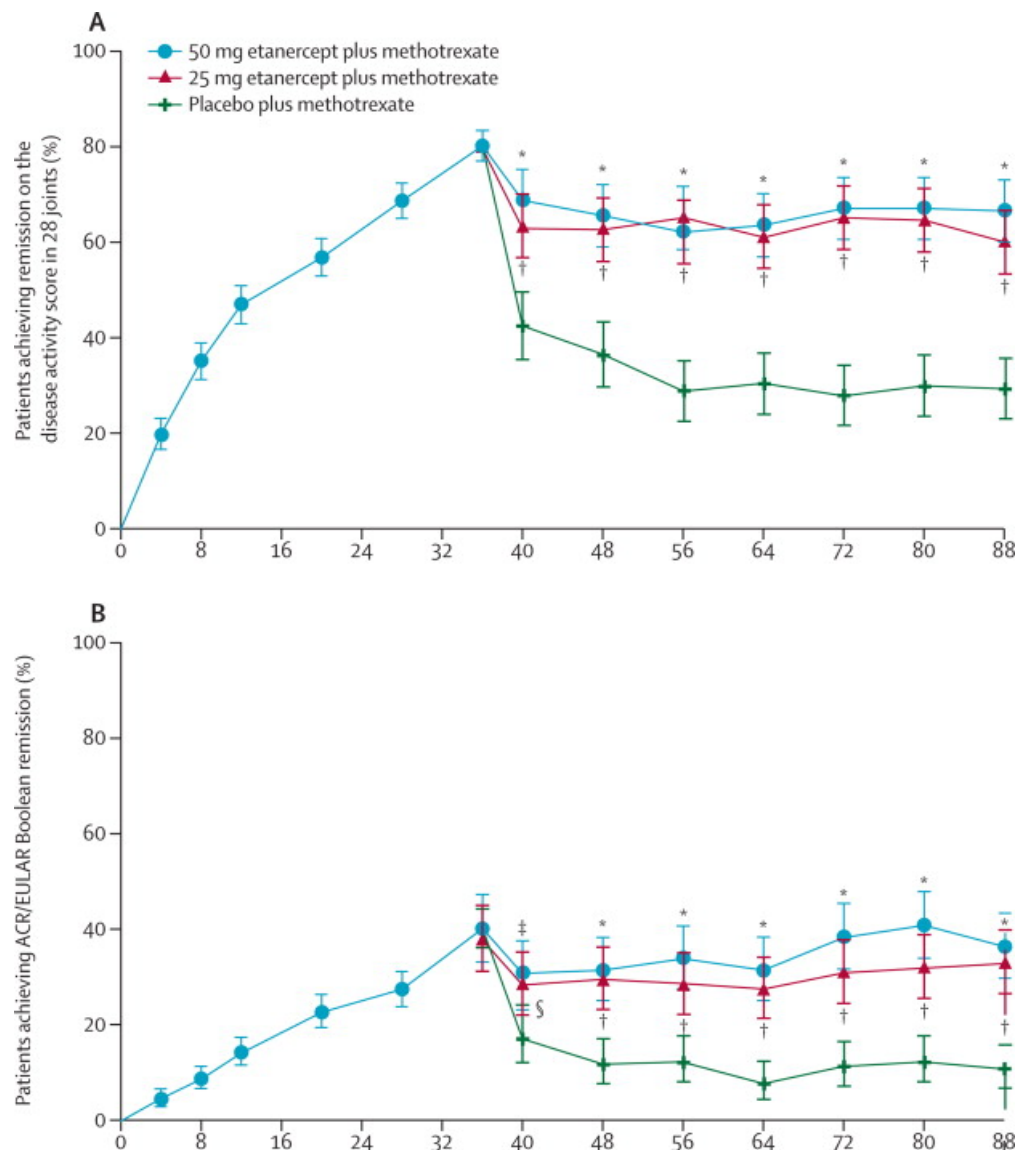
- Potential Cost Savings

- WAC for Etanercept 50mg \$3000/month
- WAC for etanercept 25mg: \$1500/month

- If you reduced dose for 1,000 RA patients, you've saved (someone) 1000pts x \$1500/pt/month x 12 months = \$18,000,000

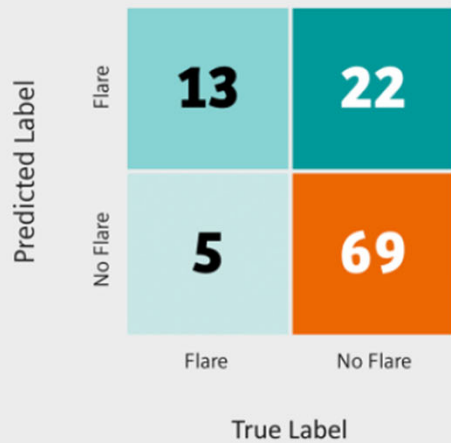
- What could we do with this \$\$\$?
- Could we expand patient access?
- Should we eliminate all copays?
- Could we eliminate prior authorizations (PAs) for all RA patients?
- Is it unethical for rheumatologists to share in this savings?

WAC = Wholesale acquisition cost



# Predicting Flare Risk: Rheums +- AI

Confusion matrix of flare prediction



(A)

- **Flare Prediction Model Accuracy**

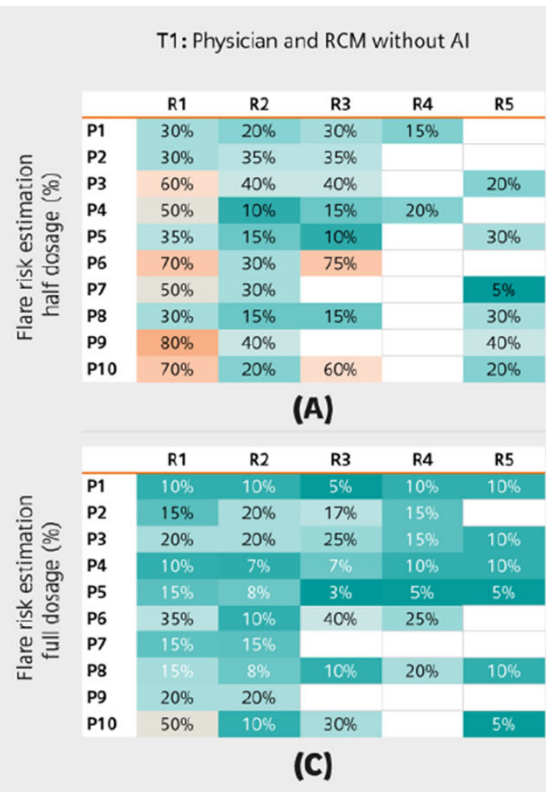
- Se=0.72; Sp=0.76
- PPV=0.37; NPV=0.93

- Note: n=113 patients

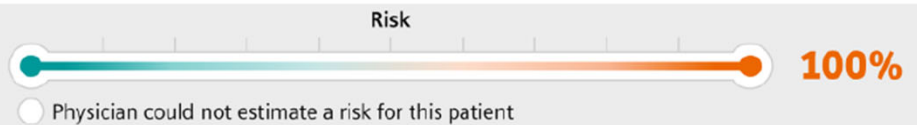
- AI model includes (only) 10 variables

True Positive Rate

1.0  
0.8  
0.6  
0.4  
0.2  
0.0



0%



# Will AI Put Me Out of a Job?

What

What AI can do...

rheumatologists do...

- Takes a patient's history

- Look people in the eyes, shake their hand

- Recommends which components of the physical exam to do

- Performs a physical exam

- Looks up the HCPCS code to bill for a joint injection

- Performs the joint injection

- Summarize ACR RA, RA-ILD guidelines; describe results from joint MRI

- Explains what the guidelines & MRI findings means for and to THIS patient

- Provides facts (sometimes correctly)

- Provides judgement, wisdom and empathy

- Predict outcomes

- Explain predicted outcomes to facilitate patient education, shared decision-making, choose treatments, triage f/u





## Key Takeaways

1. Not accept anything coming from AI as **'ground truth'**
2. Not make AI a substitute for real databases (e.g. pubmed); verify source!
3. Not put AI in high-risk settings where it can be **'wrong'** (at least, estimate accuracy of prediction)
4. Structure AI data requests such that they are:
  1. Traceable (i.e. able to be fact checked)
  2. Navigate you to information rather than provide the information (AI can even hallucinate citations)
5. Not consider anything from AI as final





## Key Takeaways (2)

6. Expect you will iterate on AI's output to refine it
7. Not assume that ChatGPT will give you the same answer each time you ask the same question
8. Be responsible to understand what AI is being used for: encourage appropriate uses, challenge misuses
9. Consider ChatGPT to be your assistant & thought partner (AI = 'Augmented intelligence', not 'Artificial Intelligence')
10. Not fear that you are replaceable

# Summary & Conclusions

- Lots of hype with AI, but used judiciously, AI offers substantial promise in numerous areas of medicine including RA
- Potential Pitfalls
  - Privacy & security
  - Long-term repercussions: **who are you licensing your data to? What are they permitted to do with it? Do you trust them?**
  - Generalizability & transportability
  - Implicit bias
  - Frenzied race to avoid obsolescence -> continual evaluation
  - AI-based interventions should be rigorously tested
- AI has applications in rheumatology science, research, advocacy, medical reimbursement, and education
  - **Who pays for this technology?**
  - **If cost savings are realized, who are they realized for?**
  - **How can patients (and providers) share in these benefits?**

# Acknowledgements & Collaborators

## ▪ Curtis P30 BIGDATA Digital Health Lab

- Tarun Arora, MS
- Tim Beukelman, MD MSc
- Cassie Clinton, MSHA
- Emily Holladay, MPH
- Ye Liu, PhD
- Robert Matthews
- Nick McCormick
- Amy Mudano, MPH
- Alex McAnnally
- Fenglong Xie, PhD
- Yujie Su, MS
- Amy Wu, MS
- Jingyi Zhang, MS

## ▪ UAB

- Maria Danila, MD MS
- Jeff Foster, MPH
- Alfredo Guzman, MSHI
- Lesley Jackson, MD
- Tapan Mehta, PhD
- John Osborne, PhD
- Kenneth Saag, MD MSc
- Andrew Trotter, MS
- James Willig, MD

Twitter: @RADoctor

Questions, Comments, Kind Heckling: [jrcurtis@uabmc.edu](mailto:jrcurtis@uabmc.edu)