### Al in Clinical Care Stanford EMED Faculty Development

Erkin Ötleş **April 2024** 







# Hello, World!

Medical Scientist Training Program Fellow MD: x2024, Engineering PhD: 2022 ML Dev & Implementation Lead

Previously:

Healthcare Data Science Manager

**Epic Engineer** 

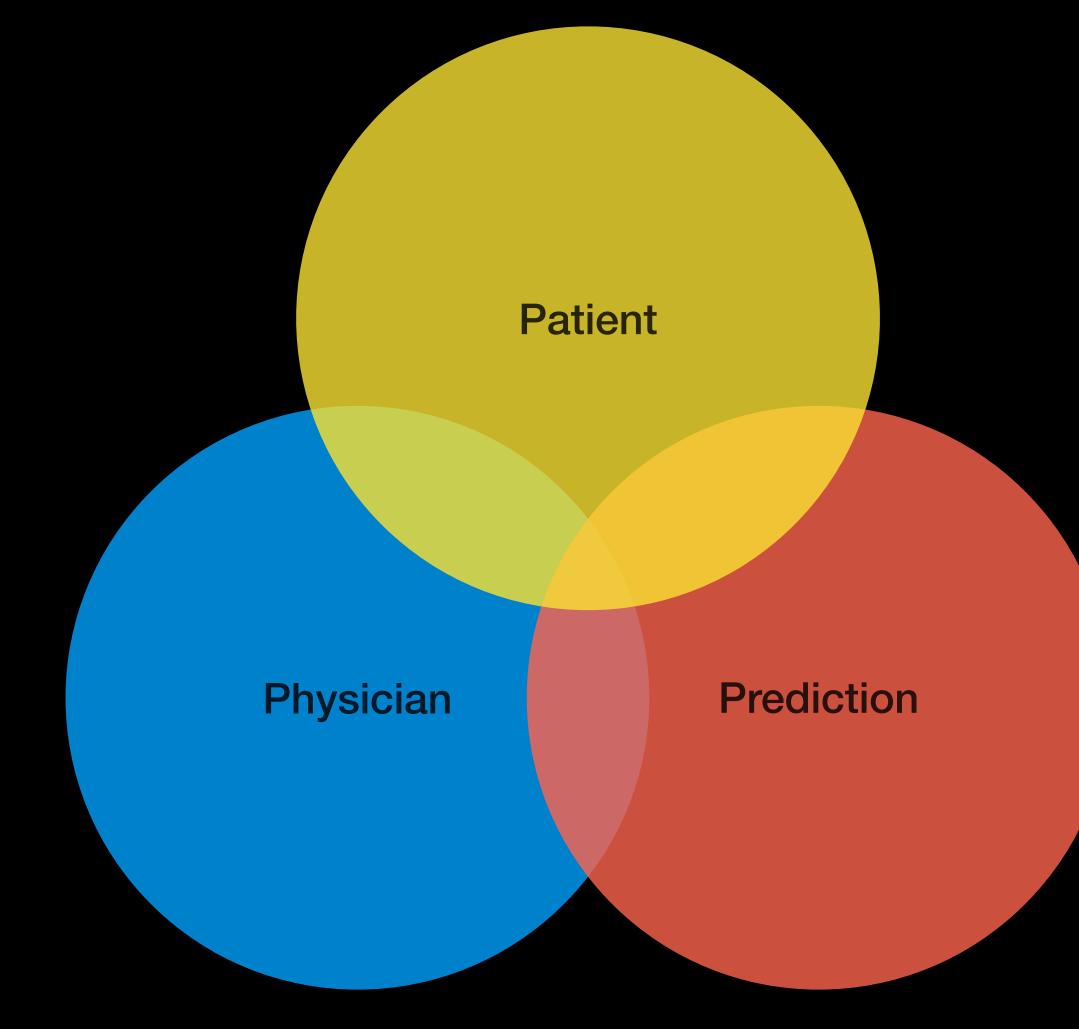




# Hello, World!

Dissertation: Machine Learning for Healthcare: Model Development and Implementation in Longitudinal Settings

Co-advised: Jenna Wiens (CS) & Brian Denton (IOE)

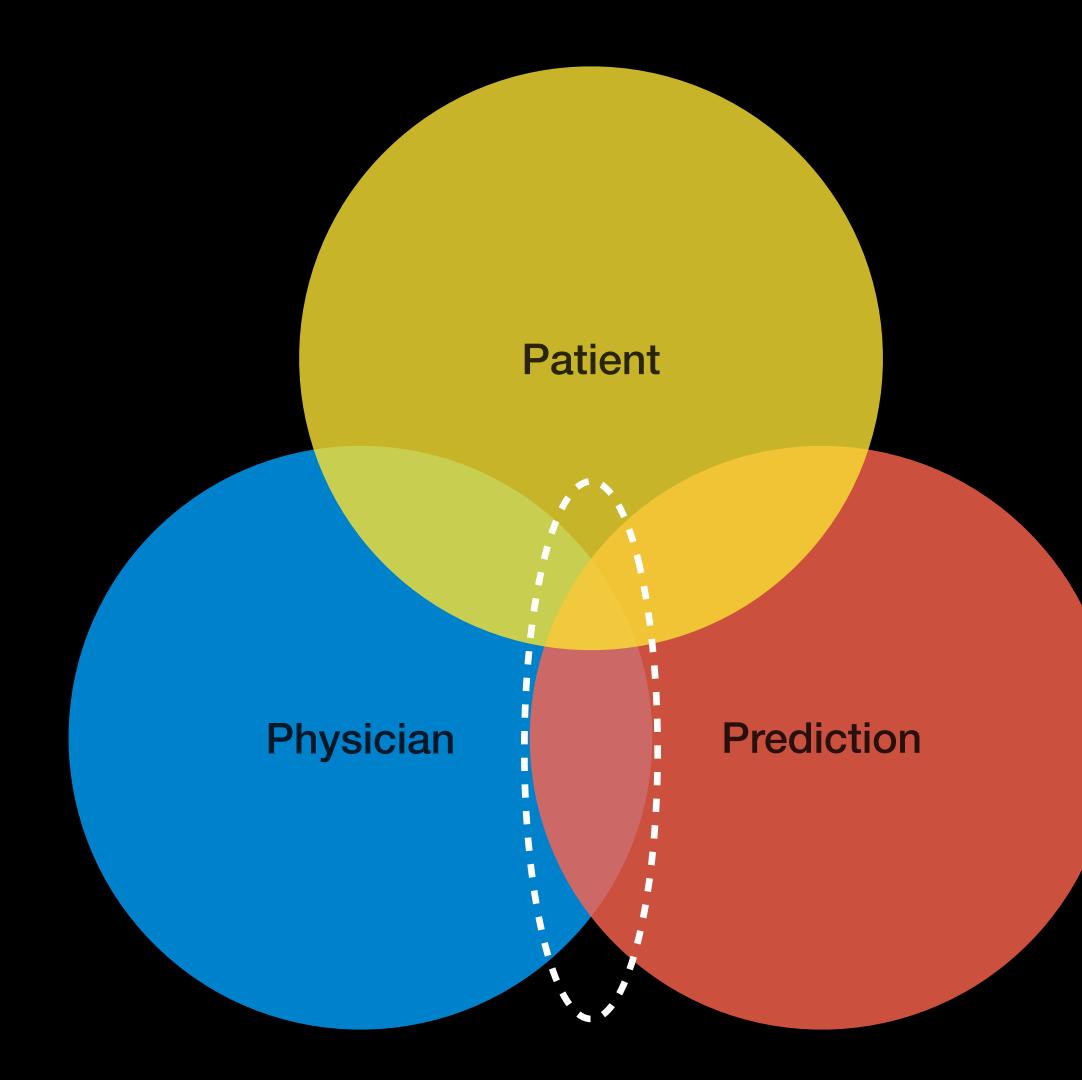




# Hello, World!

Dissertation: Machine Learning for Healthcare: Model Development and Implementation in Longitudinal Settings

Interested in computational approaches to make AI tools more useful for physicians and patients.





### Potential Conflicts of Interest

#### Advise startups: clinical summarization using LLMs

#### Patent pending: Al prediction of health outcomes in patients with occupational injuries.

Small amount of stock in various technology & healthcare companies.

### Agenda

Background

Definitions

Connections between generative & predictive AI

**Clinical Al** 

Repurposing predictive AI to do generative tasks

Constant evaluation is fundamental

Upcoming Generative Clinical AI

AI Scribes & Summarization

Medical foundation models

Discussion

Background

### First, some definitions

### inferring information) demonstrated by machines

Machine Learning (ML): field of inquiry devoted to understanding and building methods that *learn* (use data to improve performance on a task).

Adapted from wikipedia

Artificial Intelligence (AI): intelligence (perceiving, synthesizing, and

### Nesting and overlapping concepts

#### Artificial Intelligence

Machine Learning

Perception

NLP

Optimization

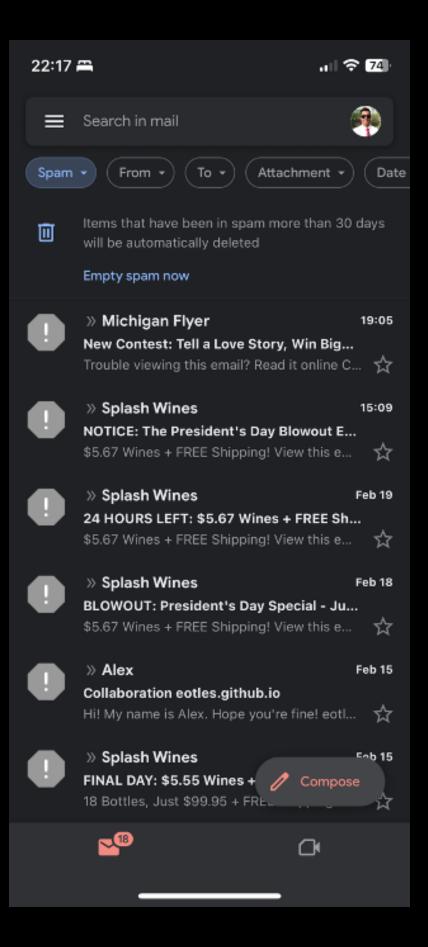
### Deep Learning

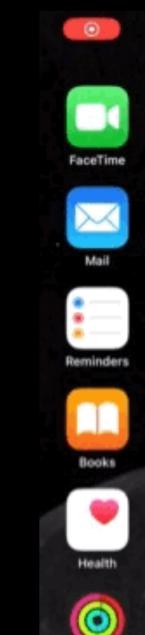
#### Statistics

KRR

HFE

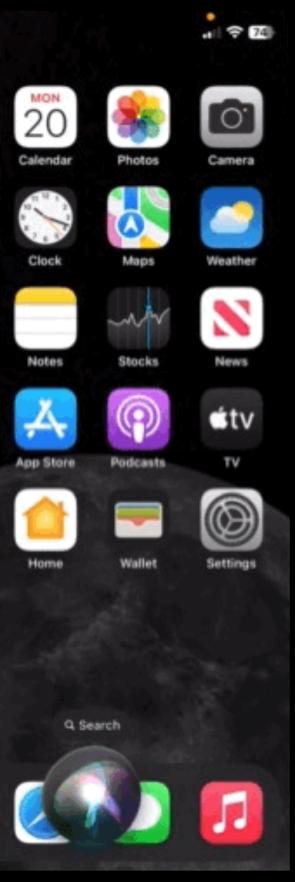
# Al is ubiquitous in everyday life

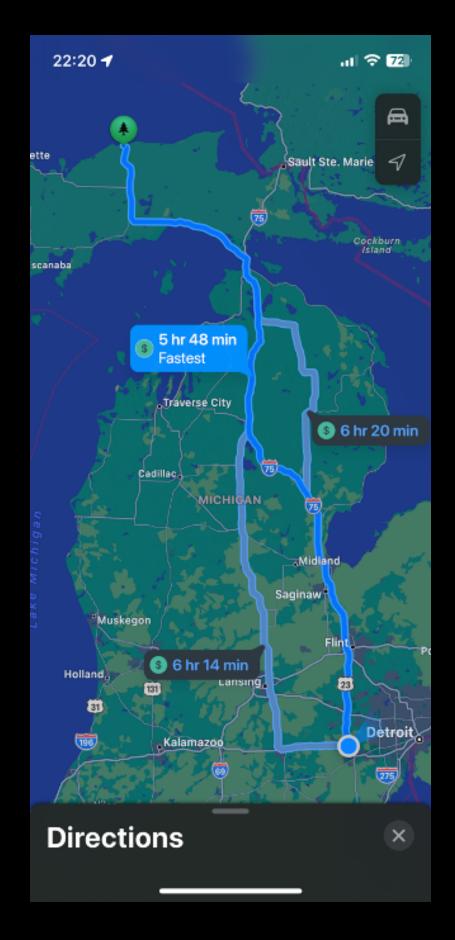


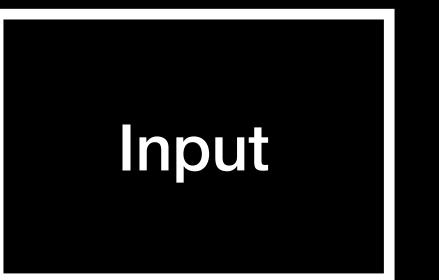


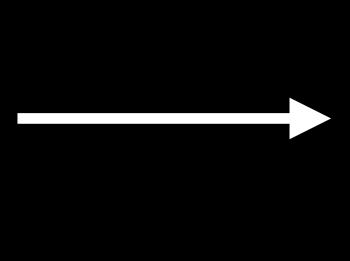














#### what we know

Age **Heart Rate** Comorbidities **Medications** 

#### Rules ML Model **Al Algorithm**

#### Model

## Output

#### prediction

**Risk of Diabetes** Length of Stay

### Al has the potential to advance medicine

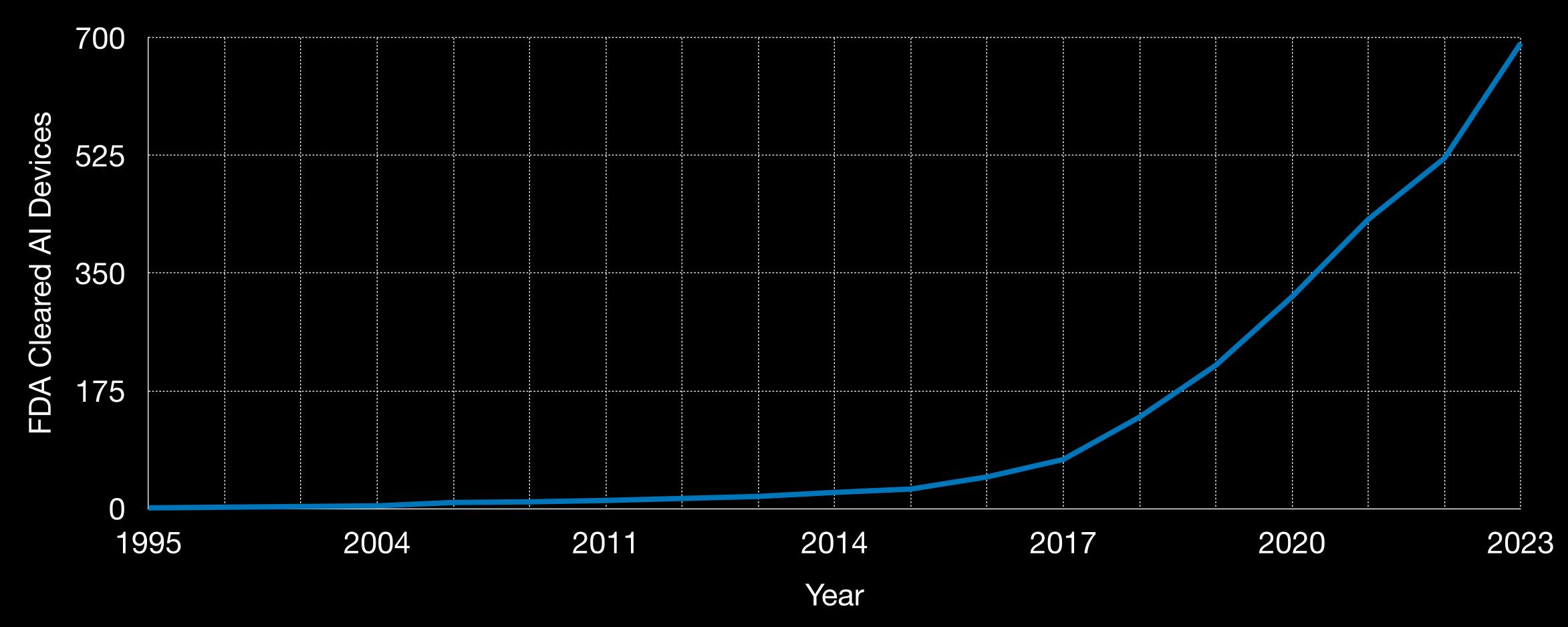


Al has techniques to rapidly **summarize** information, **predict** outcomes, and **learn** over time

Society has big expectations for AI in medicine

# Increasing prevalence of medical Al

### no FDA cleared generative AI tools as of 2023



### Generative AI definitions

**Generative AI**: AI capable of *generating data* (text, images, etc.) using generative models, often in response to prompts.

Large Language Model (LLM): language model able to capable of general-purpose language generation and other language tasks.

**Foundation Model**: a model that is trained on broad data such that it can be applied across a *wide range of use cases*.

Adapted from wikipedia

## **Connection between Al types**

#### Techniques

Optimization Simulation

Machine Learning Statistical Modeling

**Data Visualization** Data Summarization



Prescriptive

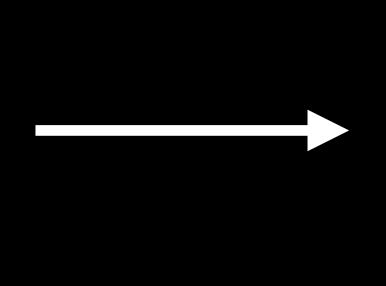
**Predictive** Α

Generative A

**Descriptive** 

### Will it rain tomorrow?

Today's Weather



what we know

It is cloudy today.



#### Model

#### Tomorrow's Weather

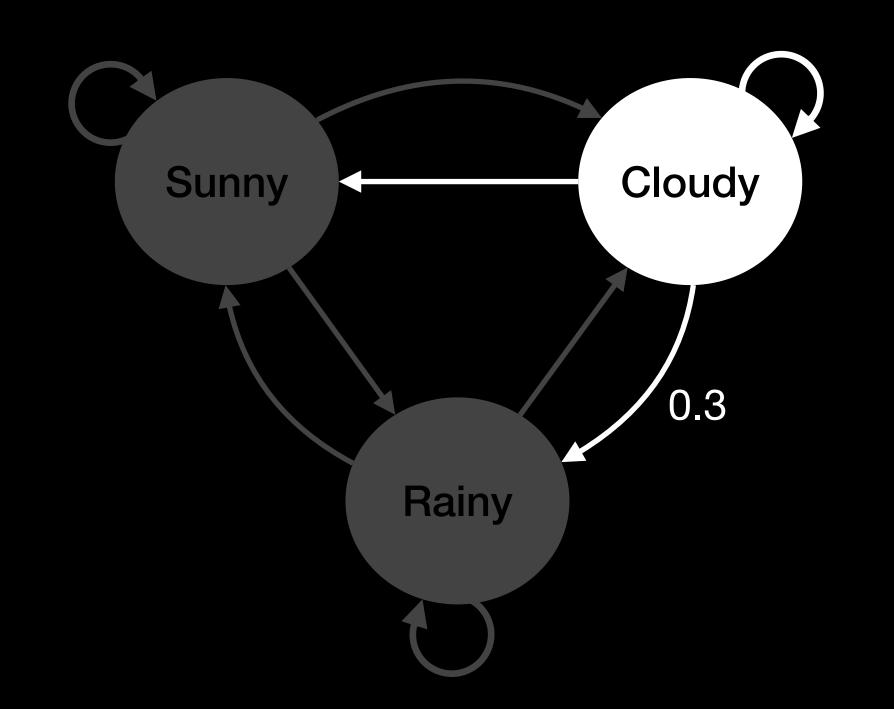
#### prediction

What's the probability it will rain tomorrow?

#### Predictive

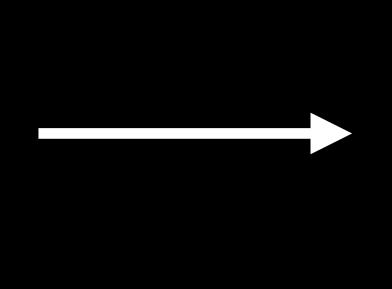
25

### Predictive use of a weather model



### What's the weather next week look like?

Today's Weather



what we know







### Tomorrow's Weather

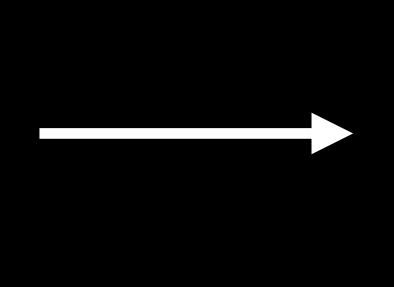
#### prediction

#### Generative

29

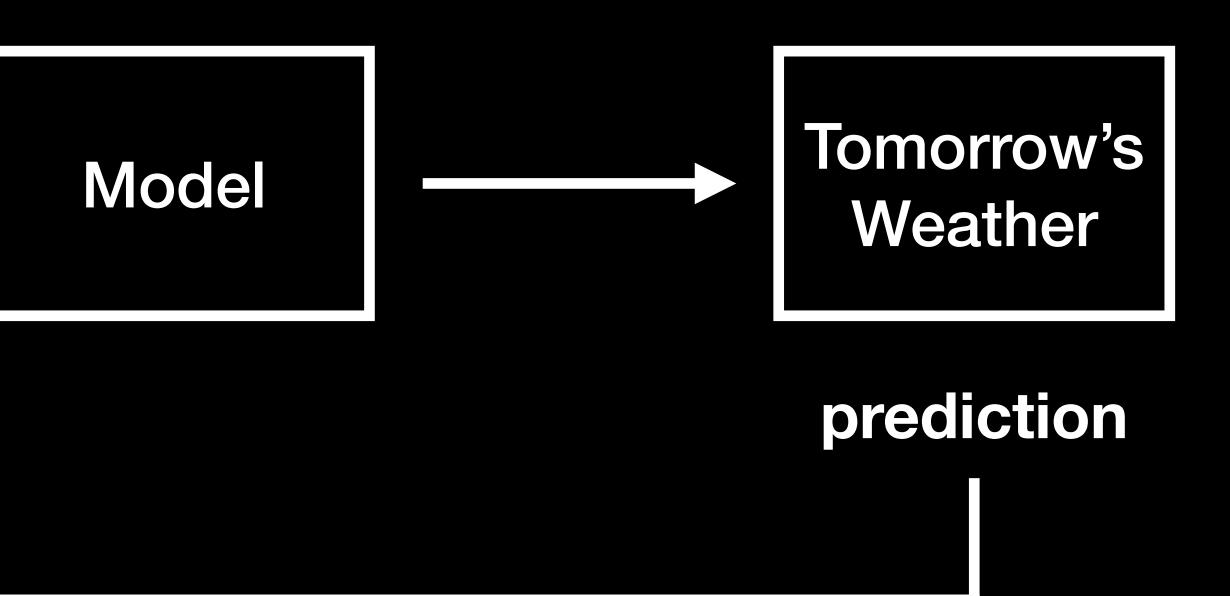
### What's the weather next week look like?

Today's Weather



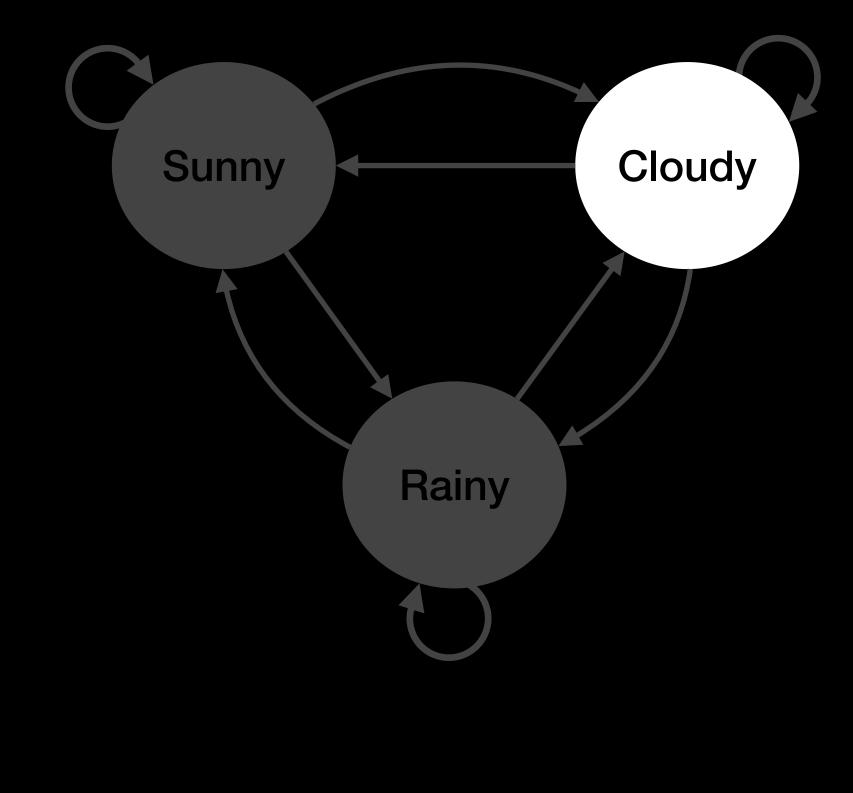




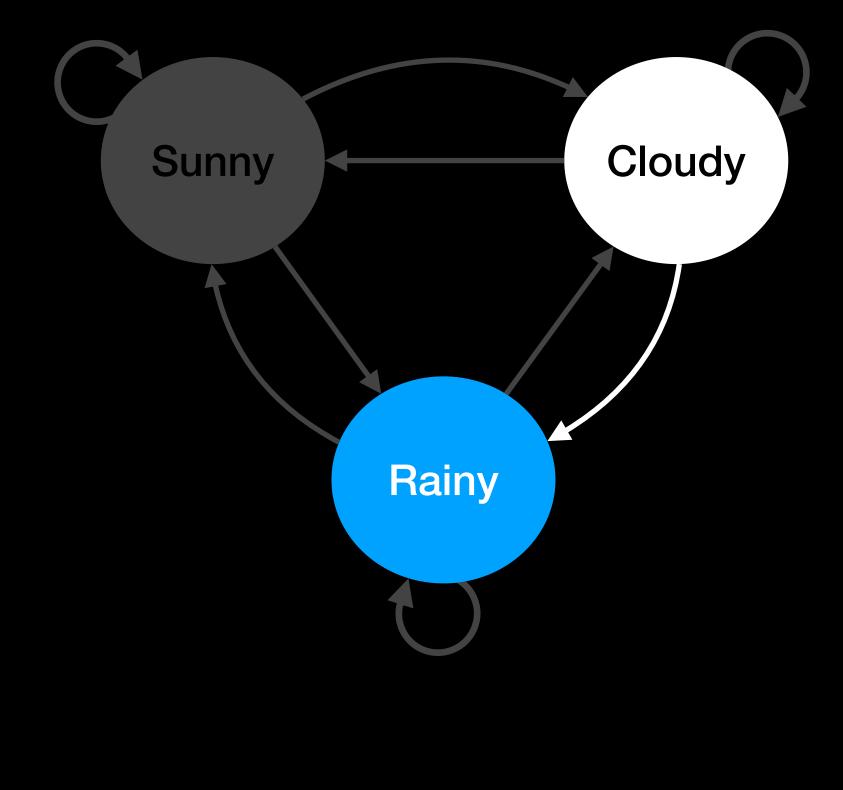


#### Generative

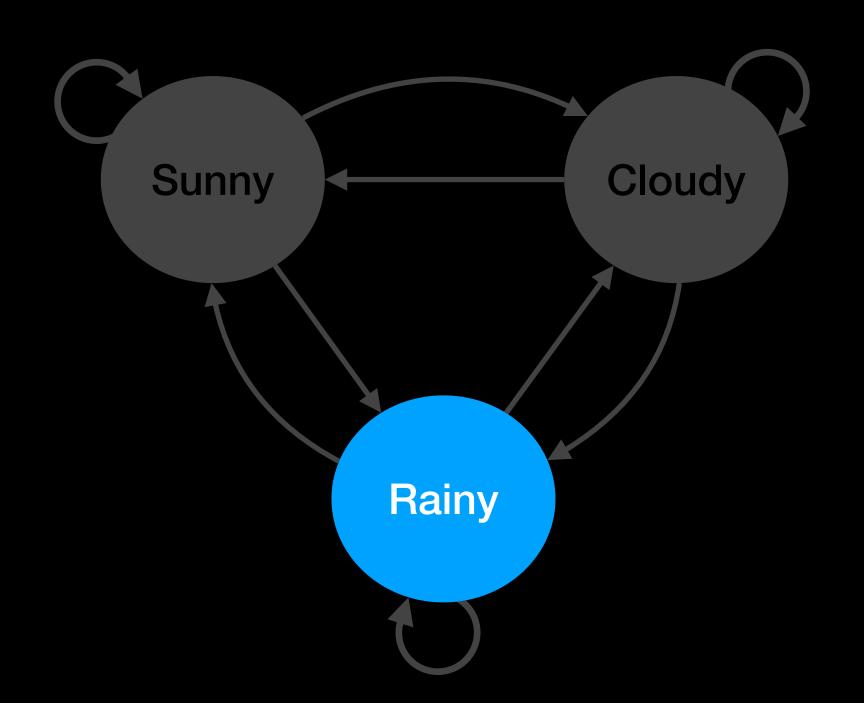
30



### Cloudy

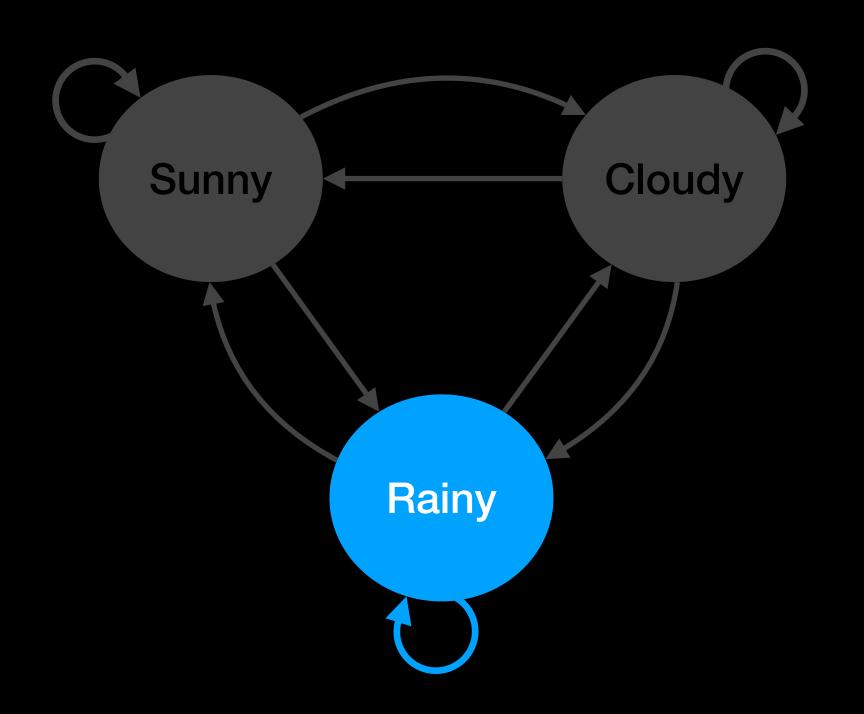


### Cloudy





#### Cloudy, Rainy





#### Cloudy, Rainy, Rainy

39

## ChatGPT = Chatbot + GPT3

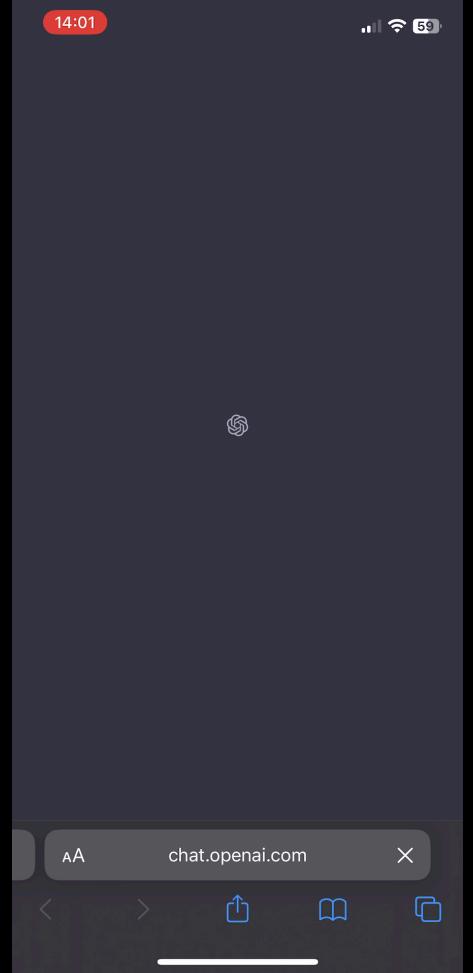
Chatbot: developed by OpenAl mix of supervised & reinforcement learning

GPT3: Generative Pre-trained Transformer 3 type of **large language model** (fancy predictive text)

"The quick brown fox jumps over the \_\_\_\_\_"

Lazy 95% Slow 2% Fun 1% ....

Trained on all available text on the internet



### Major issues with large language models

Based on what ever data it was trained on

May not be relevant, accurate, or pleasant

Generative process is inherently stochastic

Response choices and sentence construction depend on sampling distributions randomly

Hard to evaluate and verify

How often will it be right? What is right?



Clinical Al

# Return to Work

### Our OI Goal

### Predict the work-status over the course of a patient's recovery.

### Existing return to work models ignore longitudinal observations.



55 y.o. male postal worker

Gaspar 2017, MCG 2022, MDG 2022



### What is the value of longitudinal observations in return to work prediction?

observations collected beyond the time of injury?

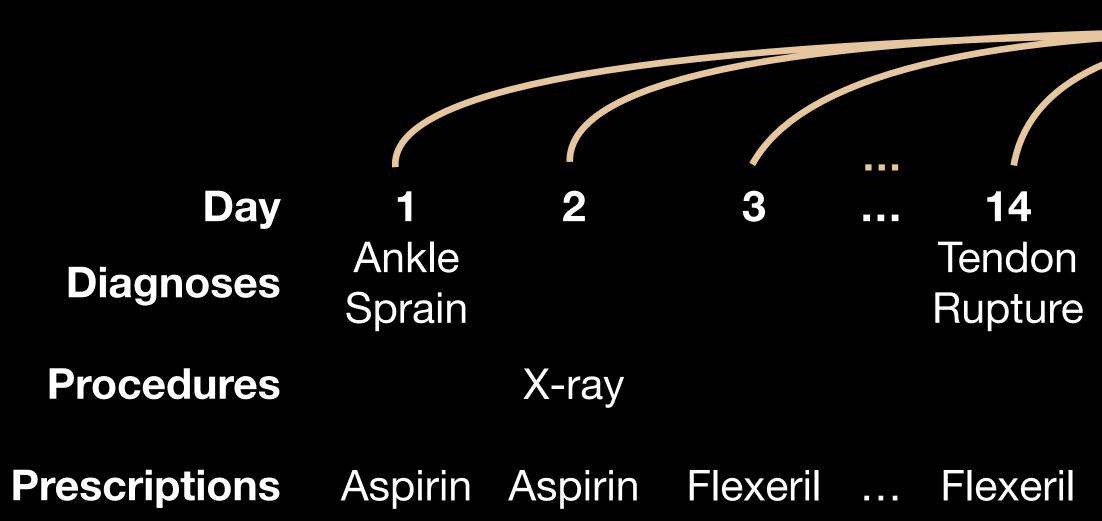
task.

- Do we observe a performance improvement when using longitudinal
- Presume longitudinal observations improve predictions in other healthcare

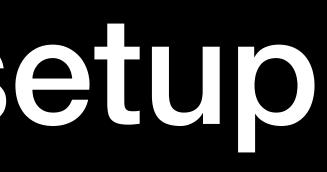
### Experimental setup

**Worker's Compensation Claims Data** 

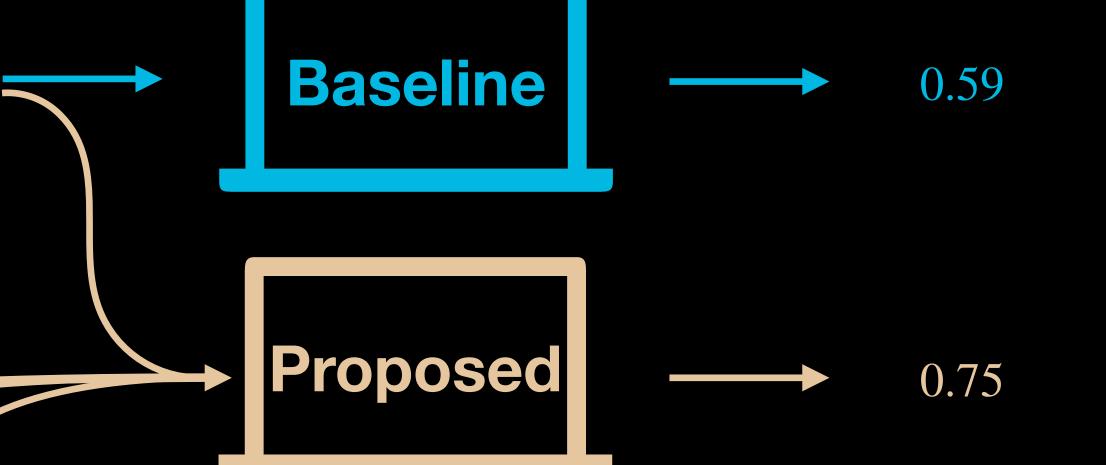




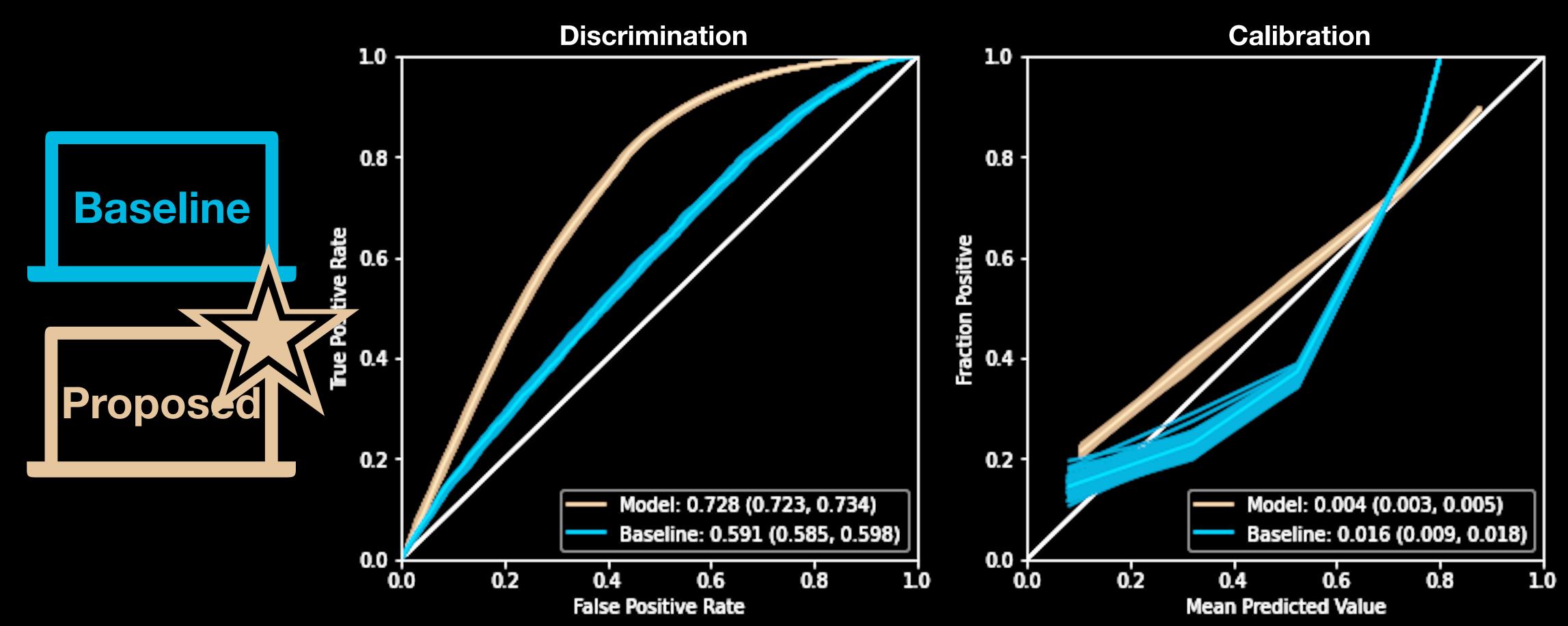
<u>Ötleş et al. 2022</u>



#### **Probability of being** at work next week



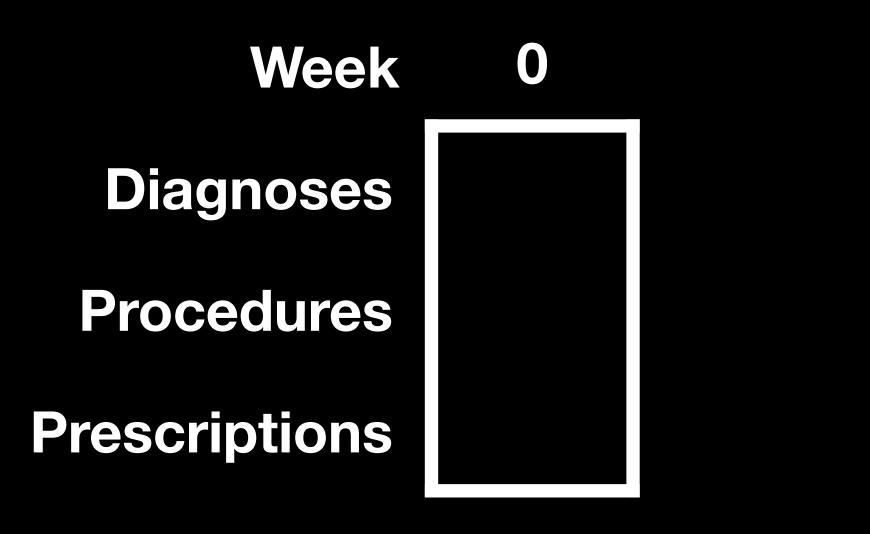
### Results



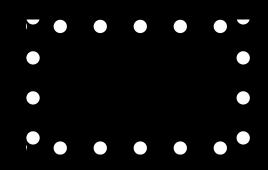
<u>Ötleş et al. 2022</u>

# Underlying predictive model can be converted to a generative model.

### Sequential prediction of recovery

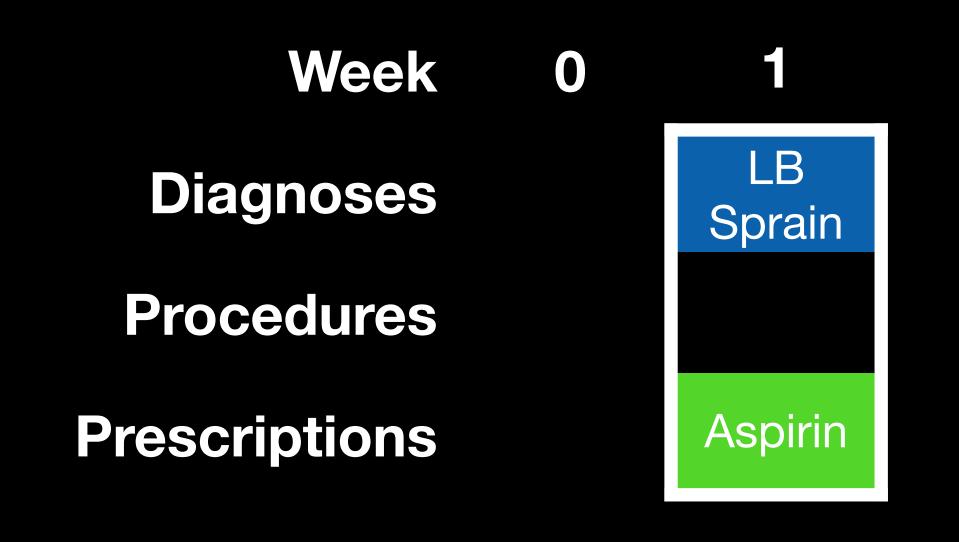






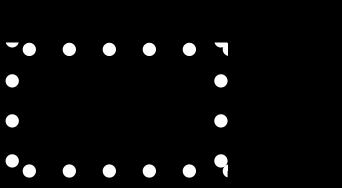
#### **25y Male Dairy Farmer** 62

### Sequential prediction of recovery





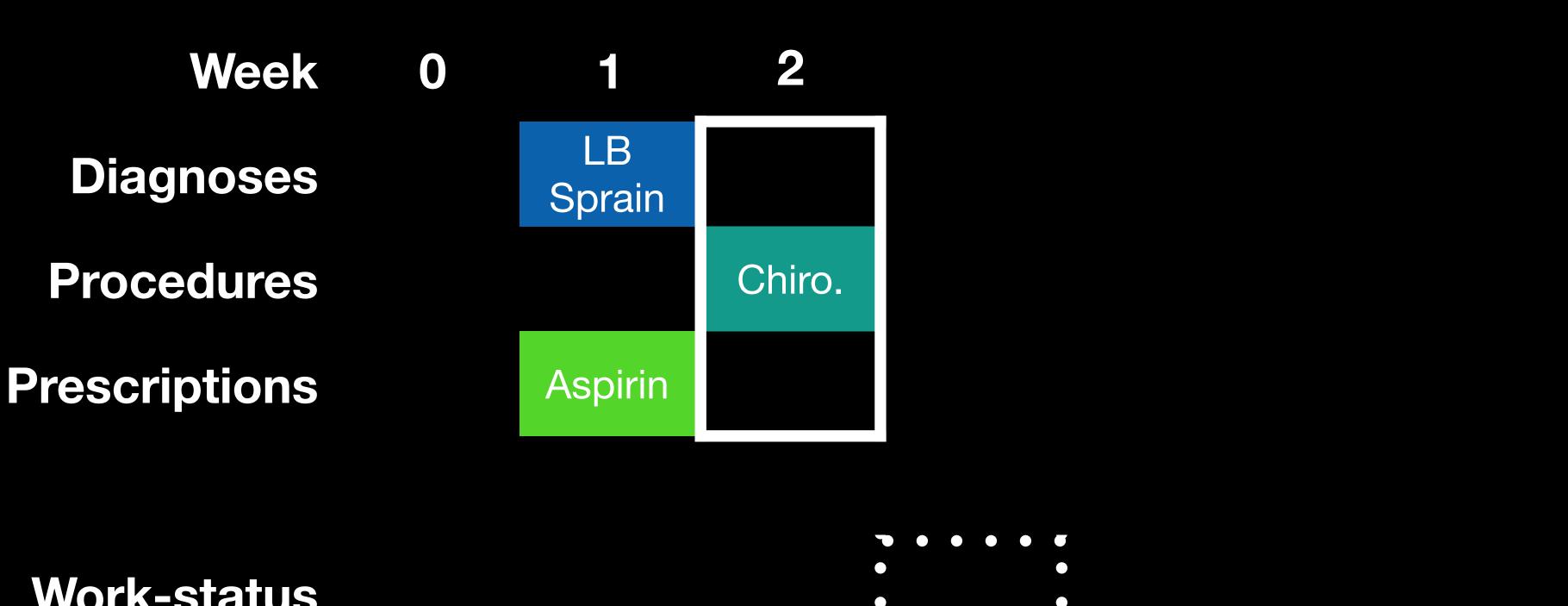




#### **25y Male Dairy Farmer**

63

### Sequential prediction of recovery

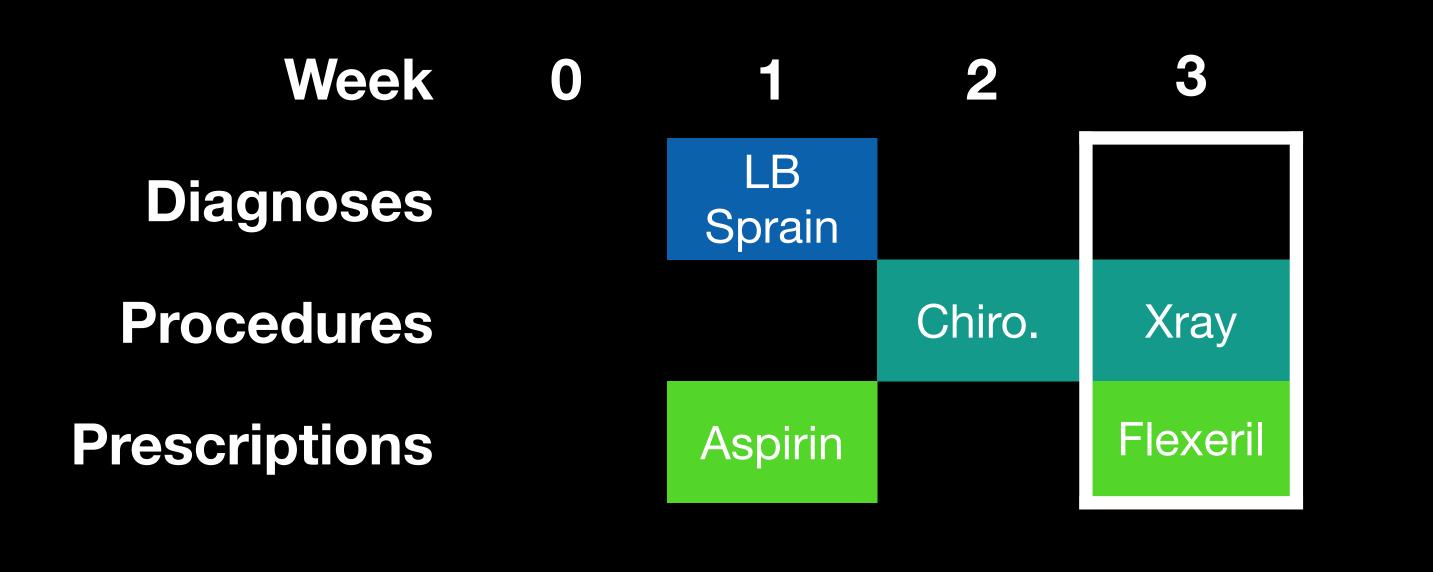


**Work-status** 



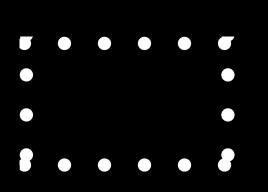
#### **25y Male Dairy Farmer**

64

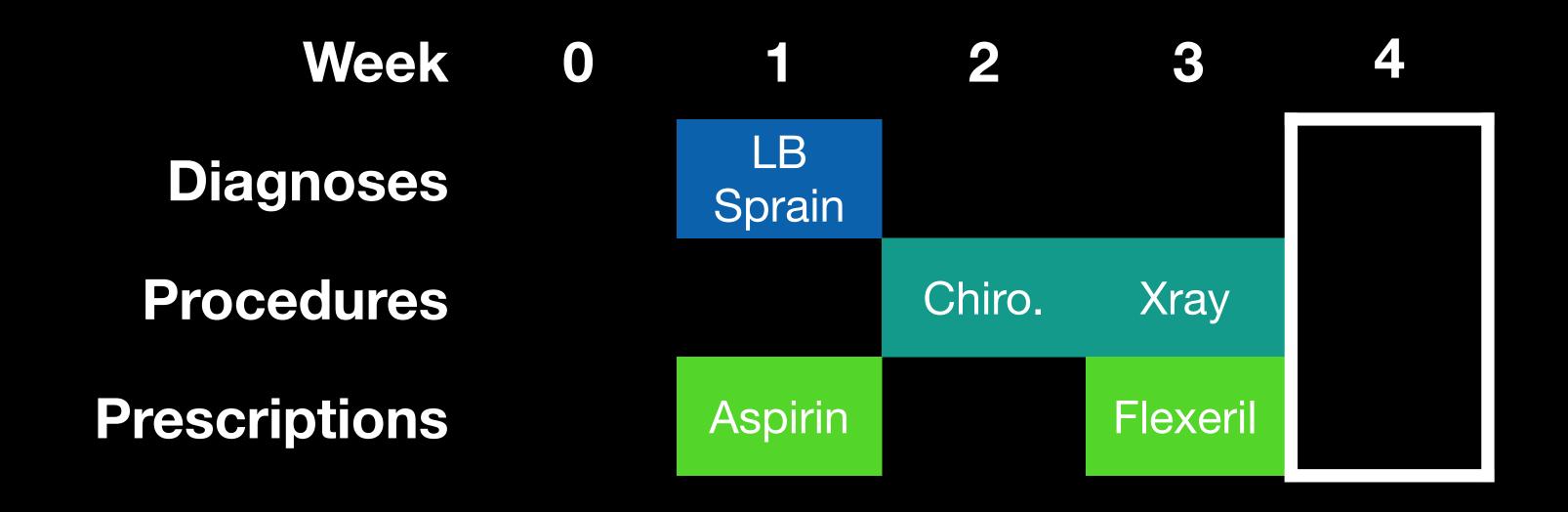


**Work-status** 





#### **25y Male Dairy Farmer**

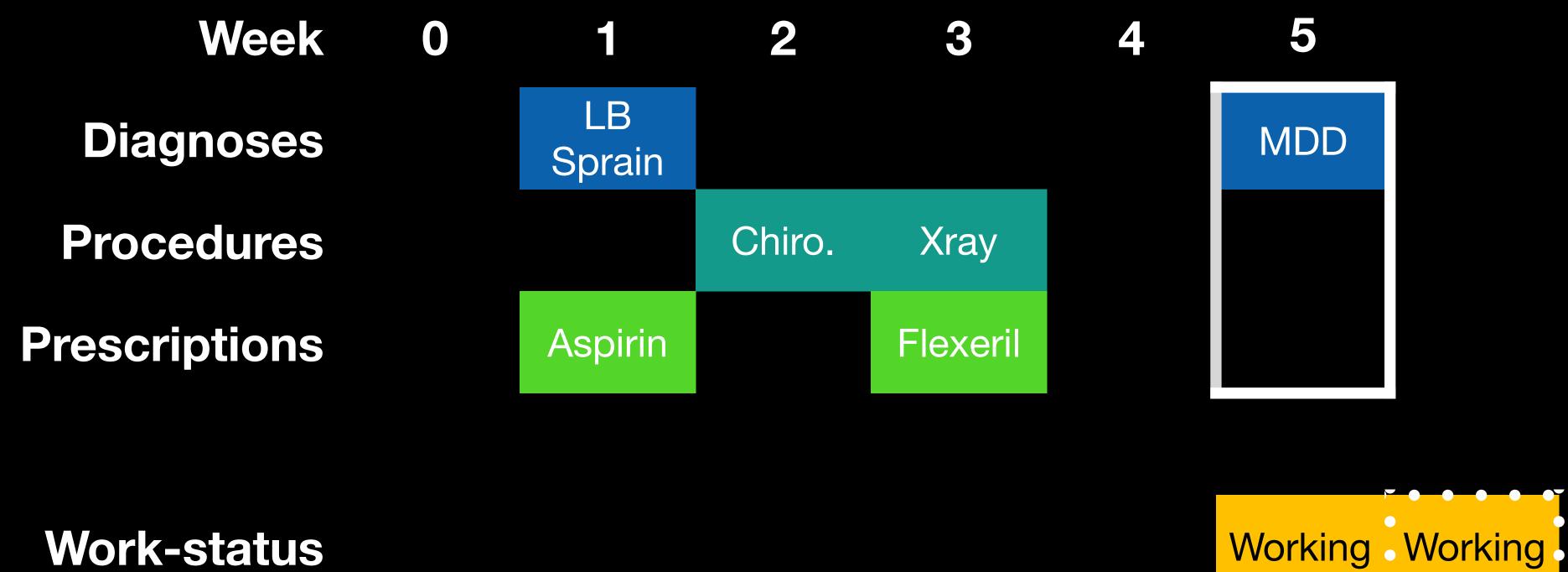


**Work-status** 



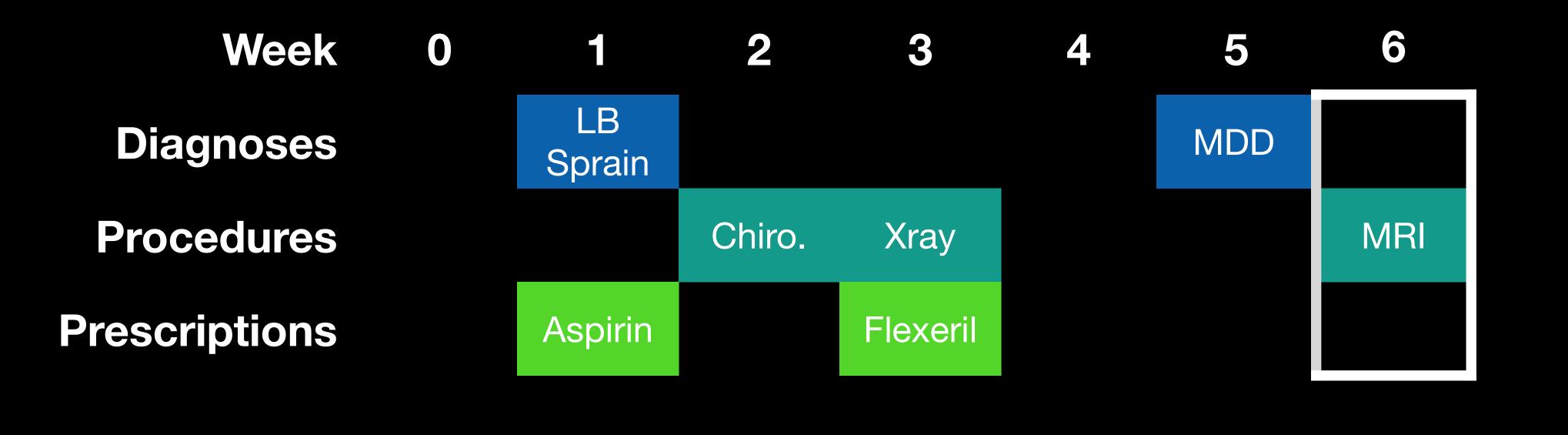


#### **25y Male Dairy Farmer** 66





#### **25y Male Dairy Farmer** 67

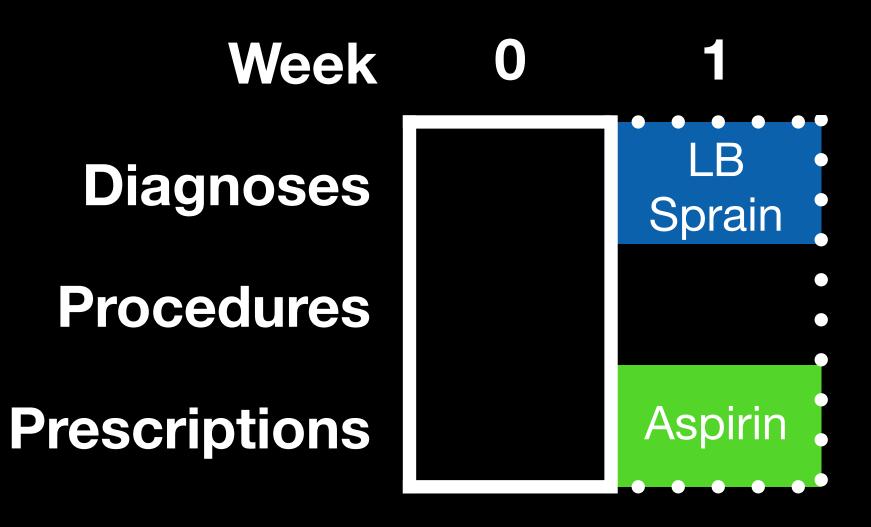


**Work-status** 



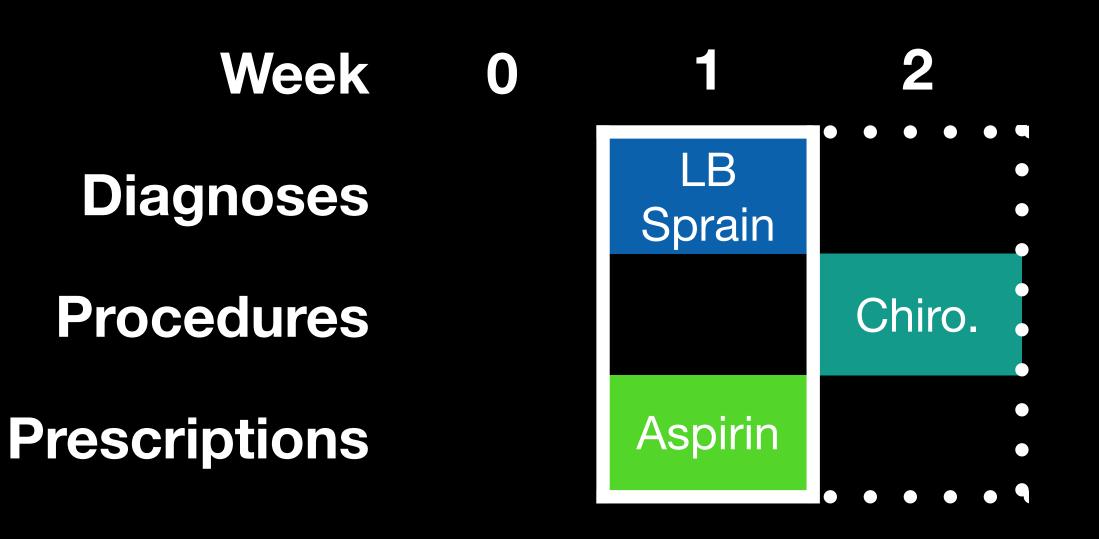


#### **25y Male Dairy Farmer**



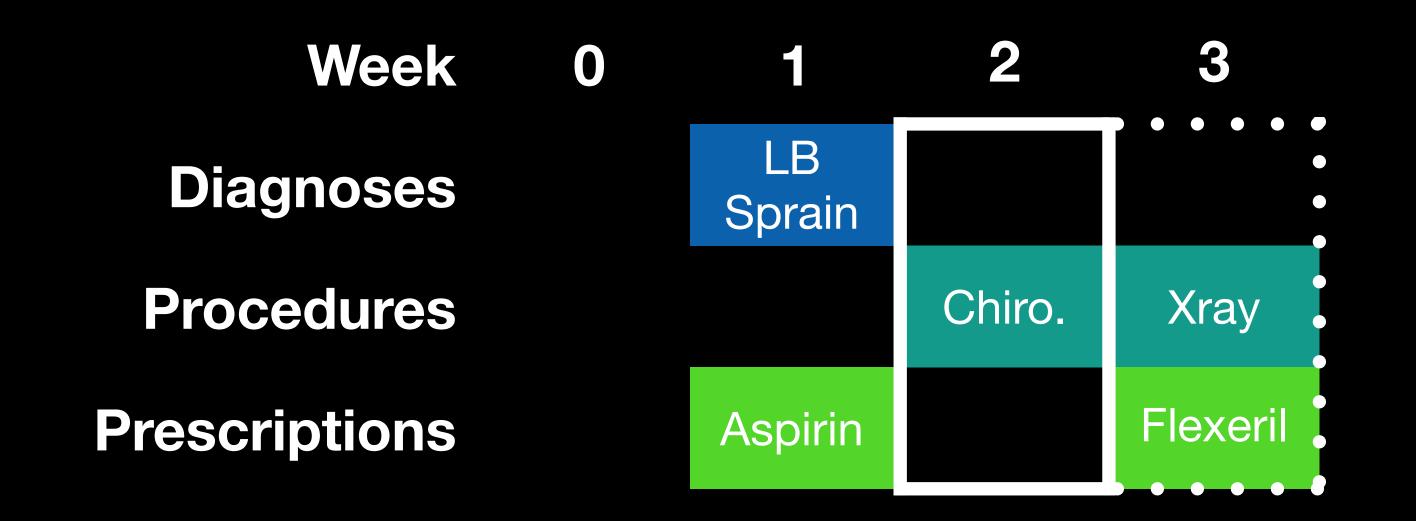


#### **25y Male Dairy Farmer**



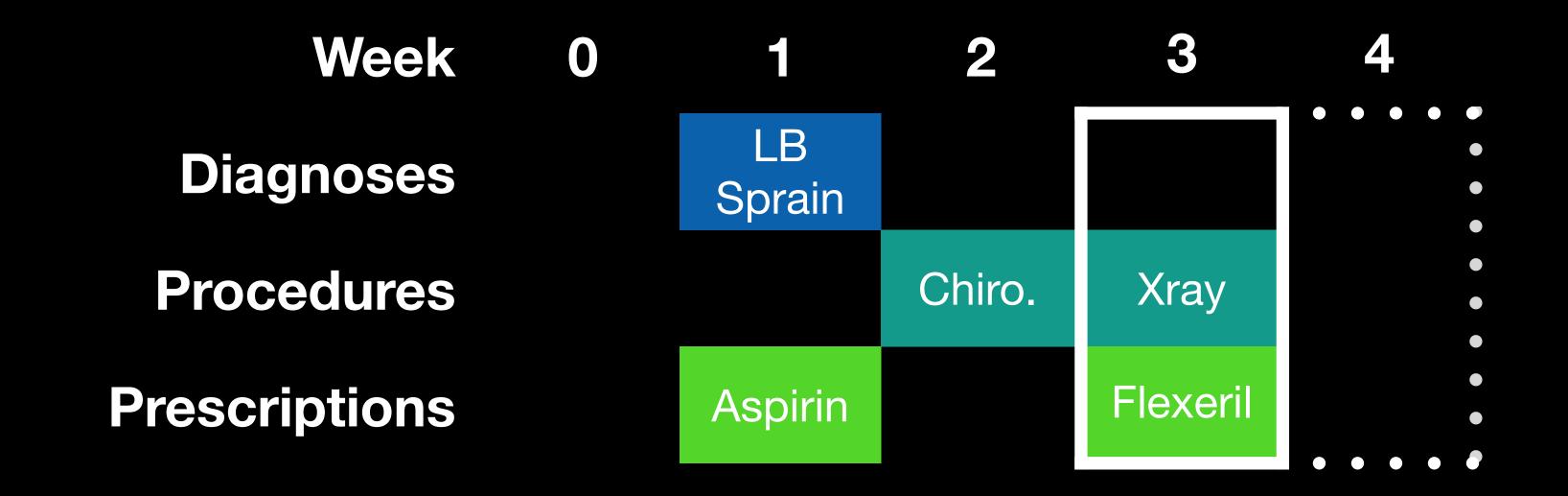


#### **25y Male Dairy Farmer**



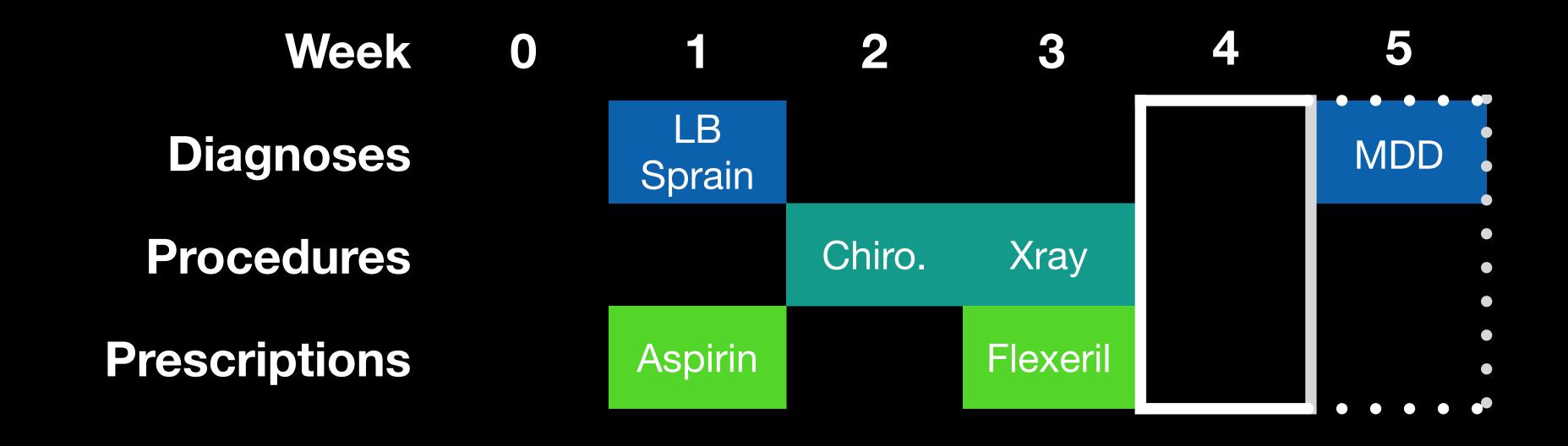


#### **25y Male Dairy Farmer**



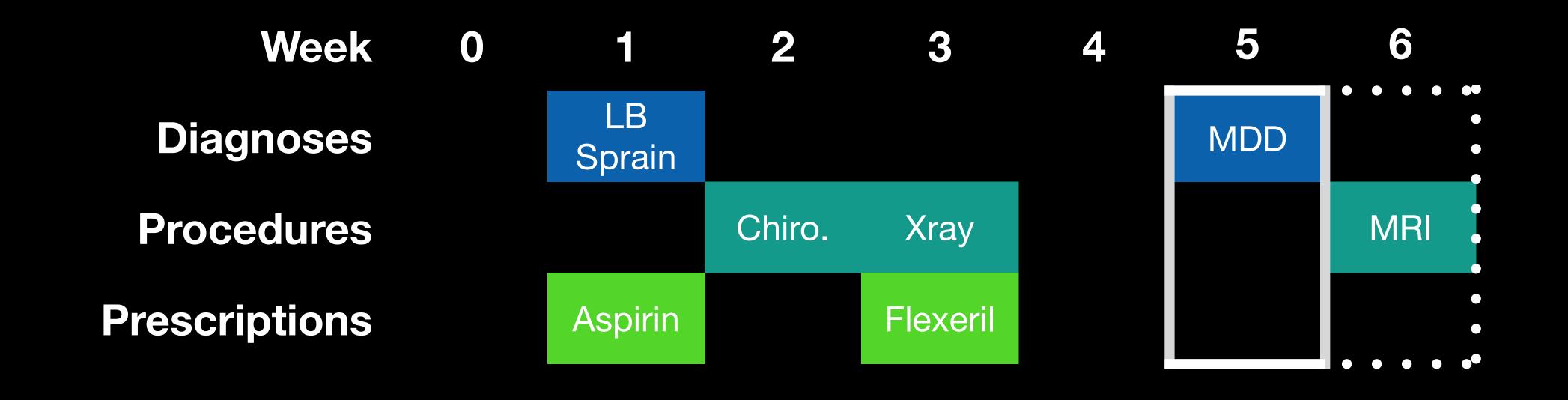


#### **25y Male Dairy Farmer**





#### **25y Male Dairy Farmer**





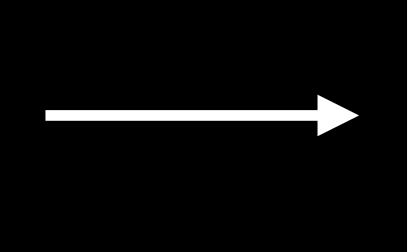
#### **25y Male Dairy Farmer**

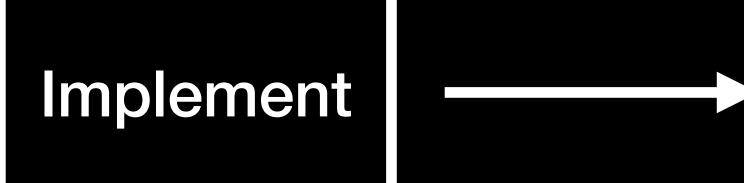
# Constant evaluation is fundamental

Prostate cancer C. difficile infection risk Sepsis

## Simplified model lifecycle

#### Develop

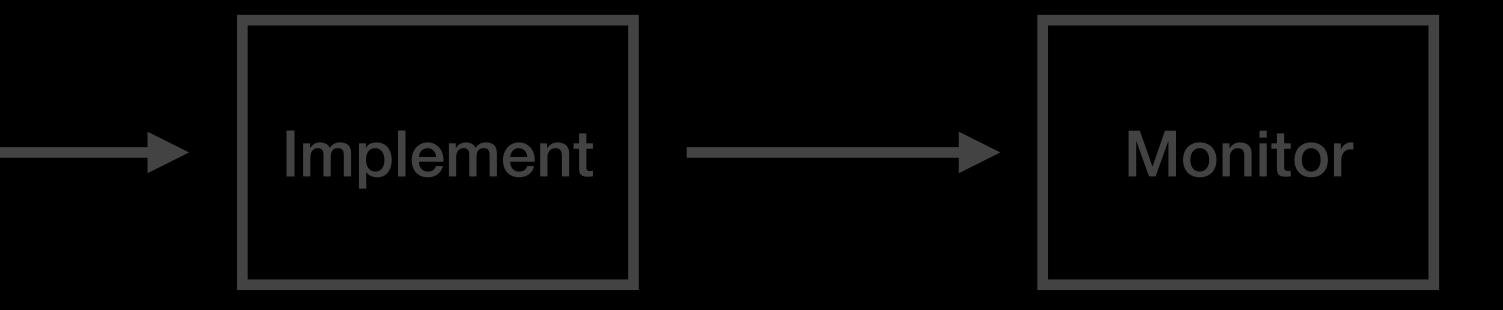






## Simplified model lifecycle



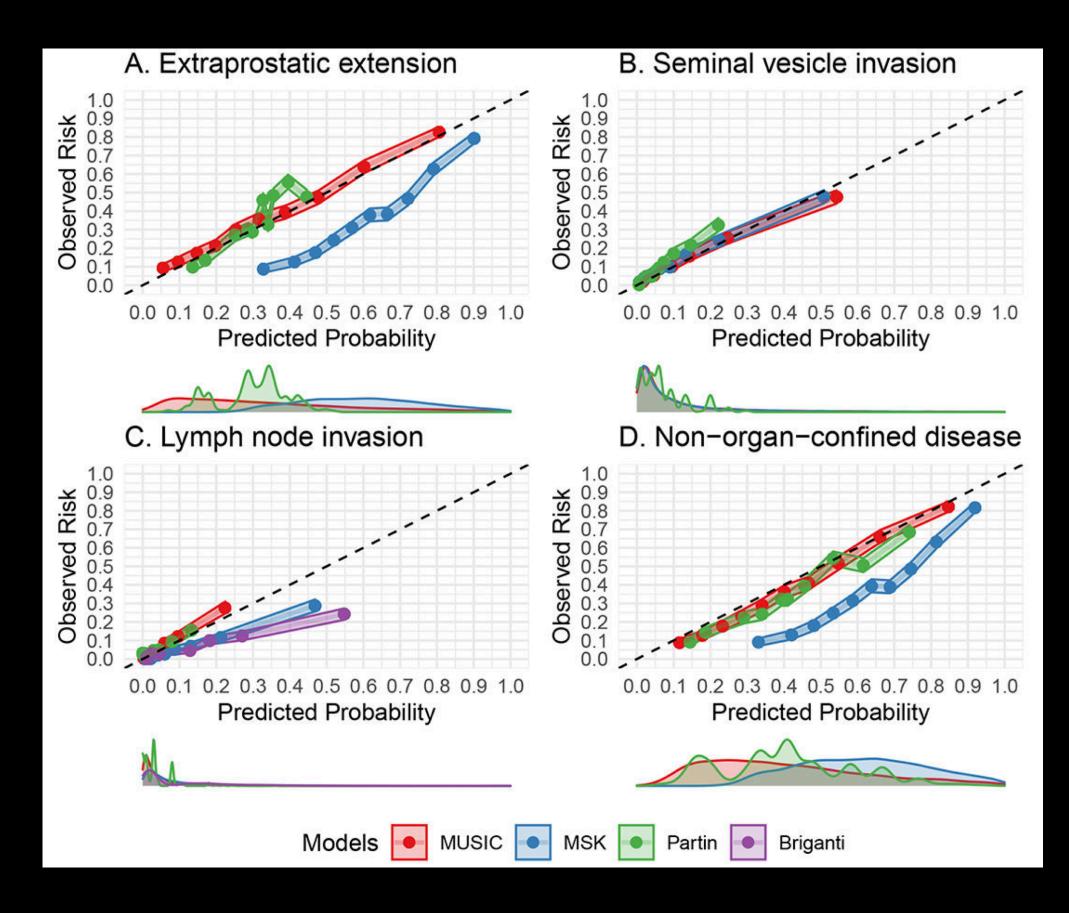


Evaluation of prostate cancer pathological outcomes prediction

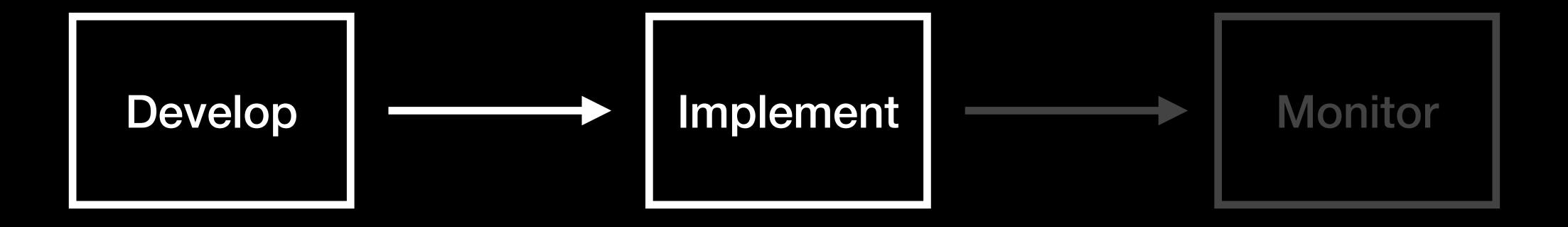
#### Internal vs. External Validation

Models developed by other institutions under perform compared to training institution specific models

<u>Ötleş et al. 2022</u>

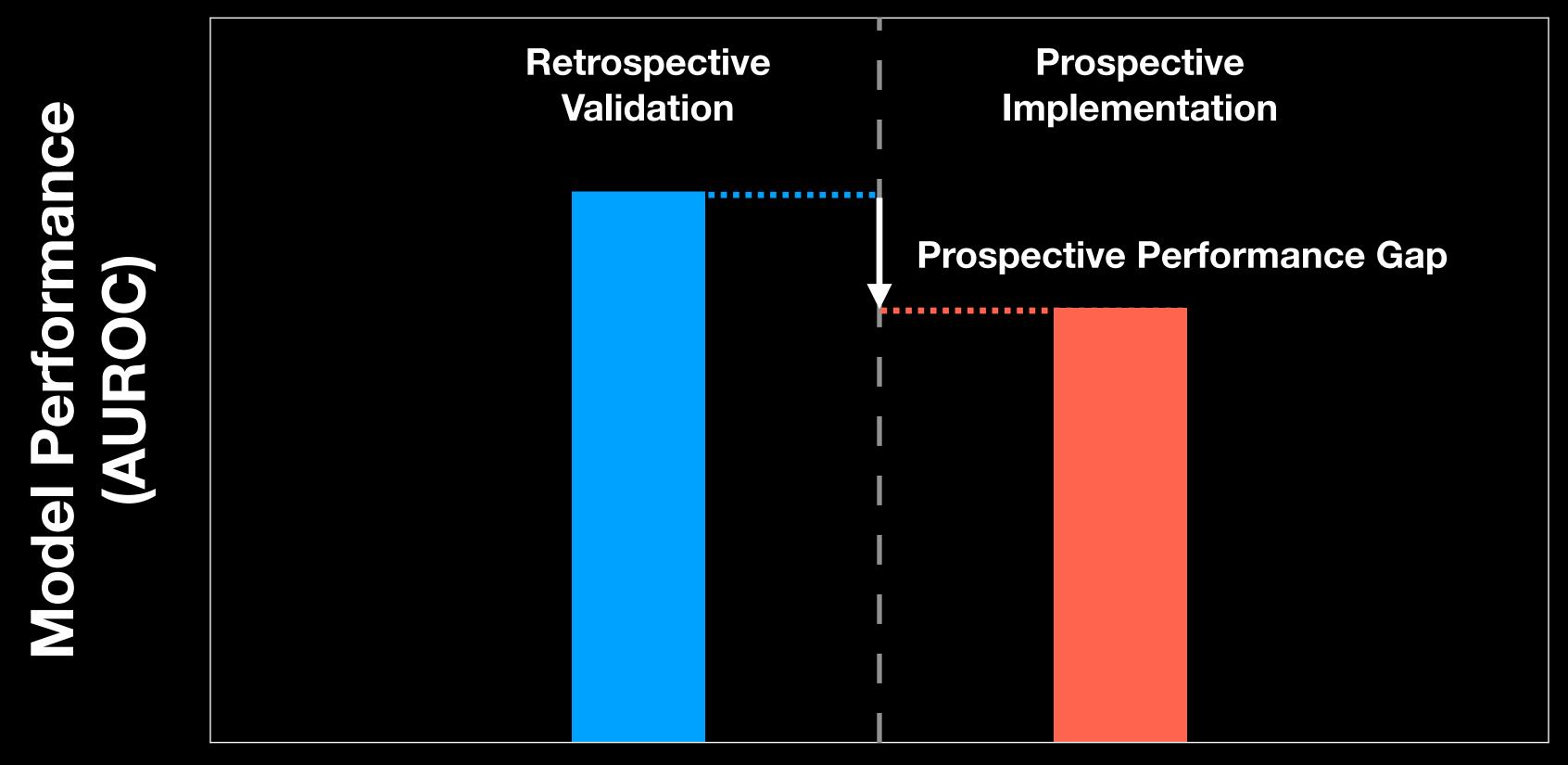


## Simplified model lifecycle



#### Prospective evaluation of inpatient C. difficile infection risk prediction

## Model performance may degrade after implementation.

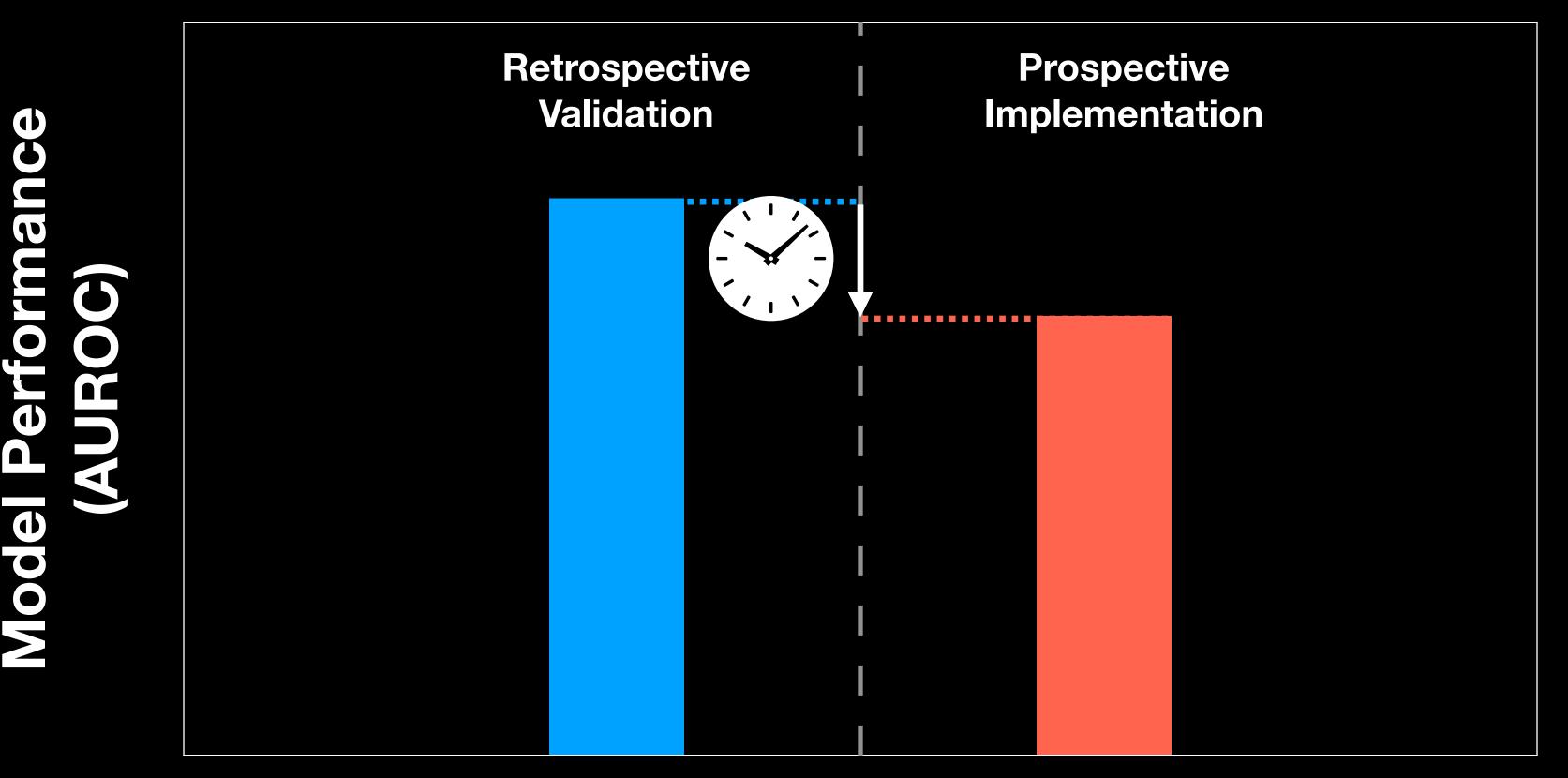


2019

Time

2020

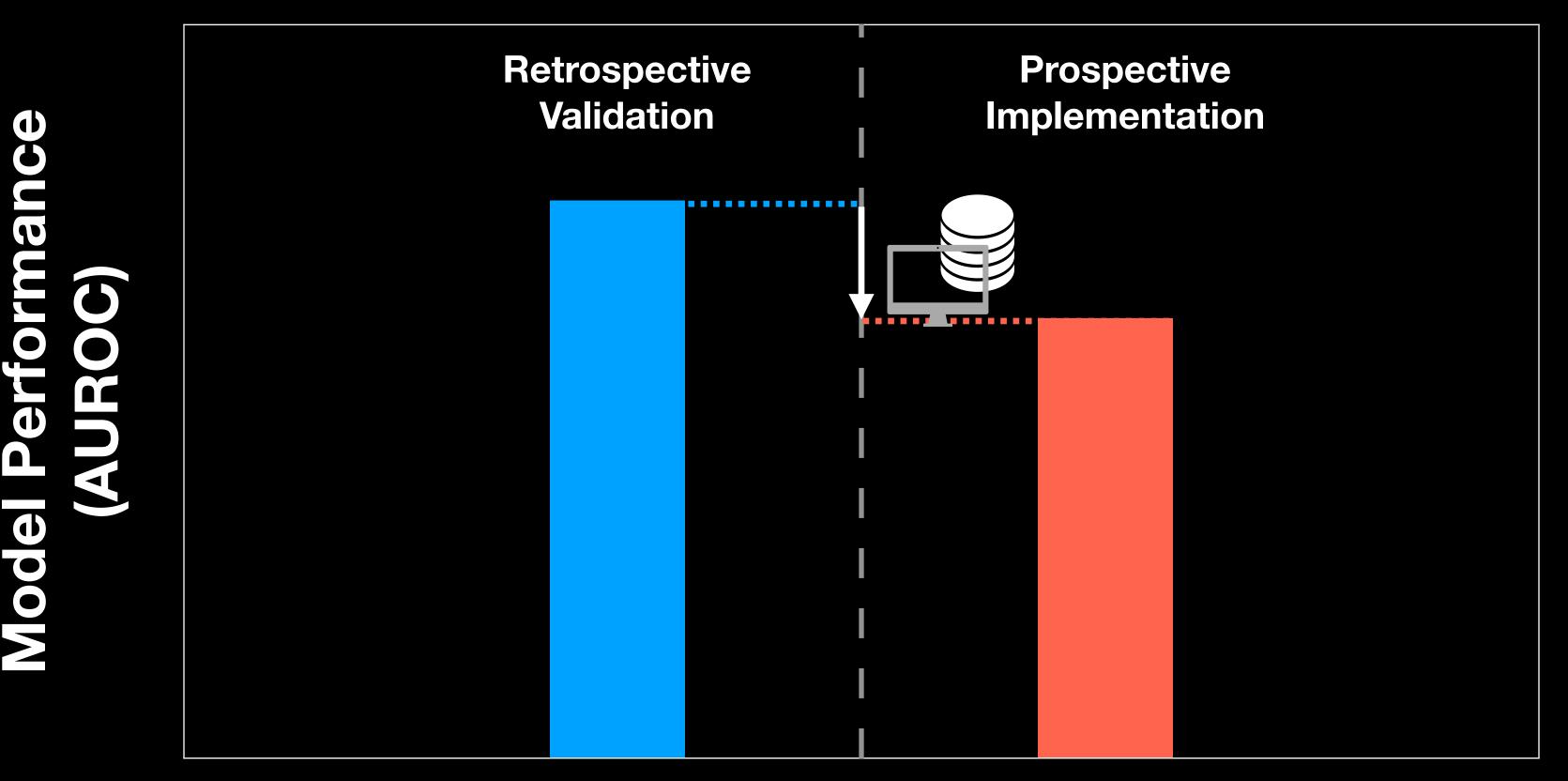
## This degradation is often attributed to changes in populations & practice that occurs over time.



2019

Time

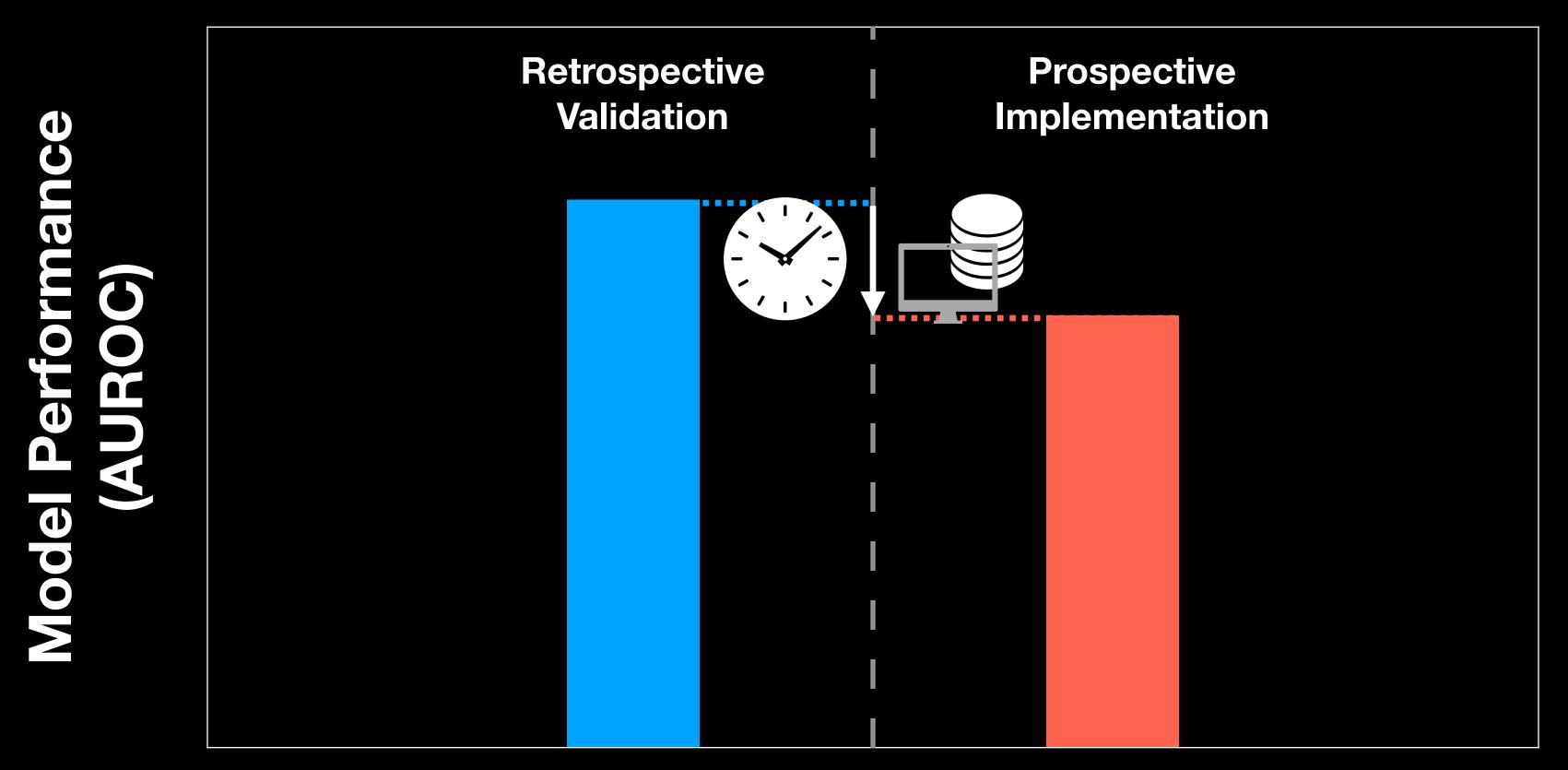
## However, changes in IT infrastructure may also affect the prospective performance gap.



2019

Time

#### Degradation due to temporal & infrastructure shift.



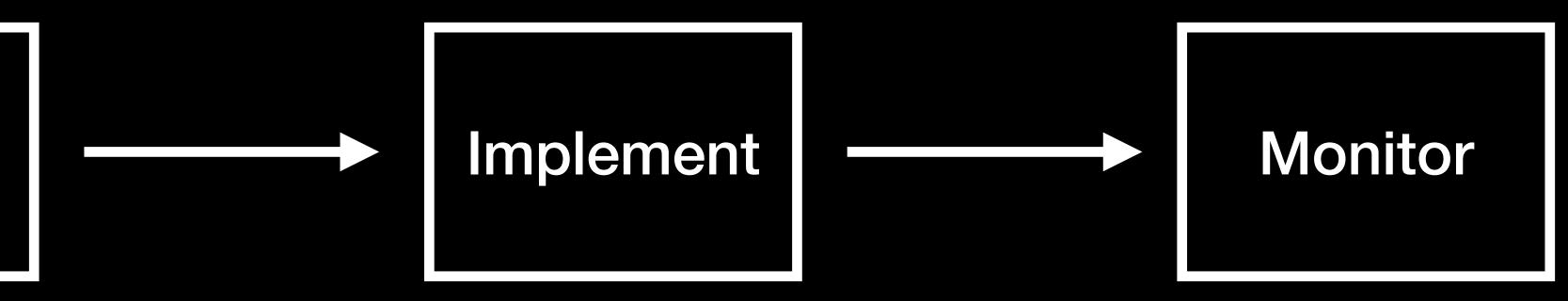
2019

<u>Ötleş & Oh et al. 2021</u>

Time

## Simplified model lifecycle

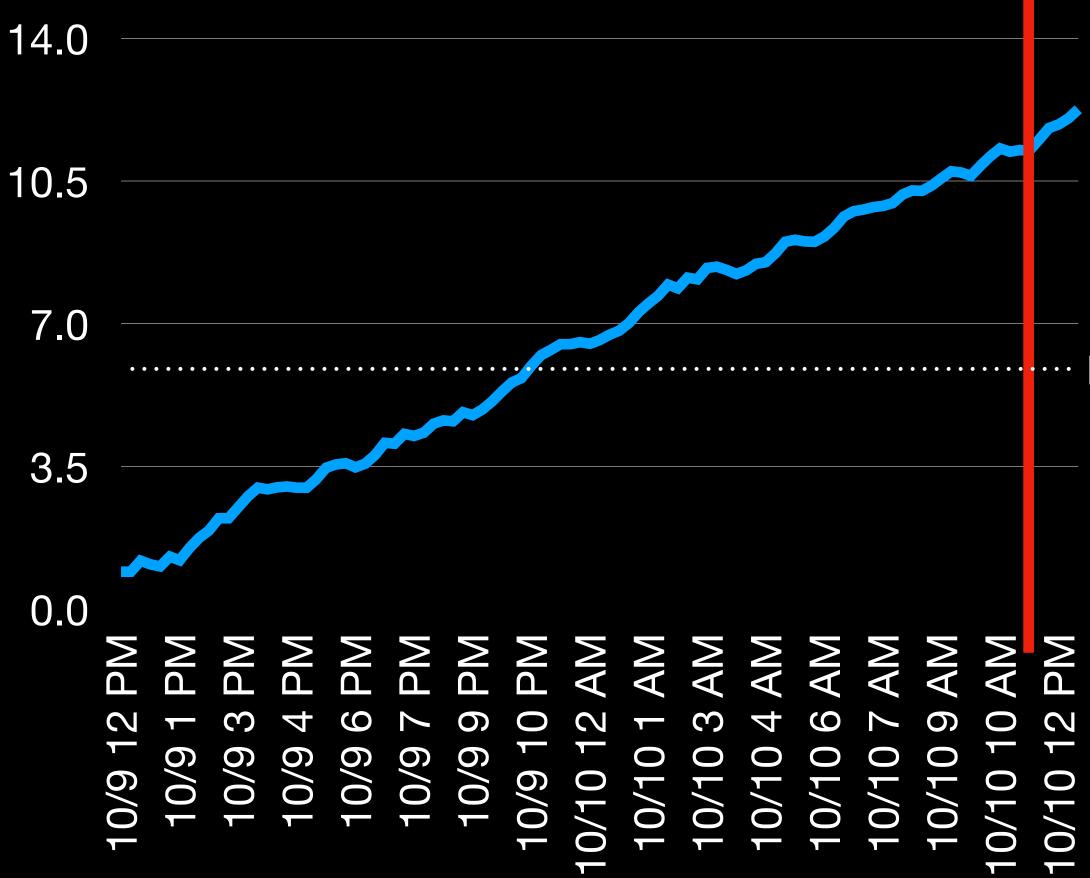




#### Prospective evaluation of Epic sepsis model

## **Epic Sepsis Model**

Sepsis!



- Development
  - Inputs: vital signs, medication orders, lab values, comorbidities, and demographic information.
- **ESM** ≥ 6
- Outputs: ICD-9 code indicating diagnosis of sepsis - timing 6hrs prior to clinical intervention
  - Implementation
    - Runs every 15 minutes on all patients in hospital
    - Expected AUROC performance ~ 0.8

#### Table 2. ESM Performance

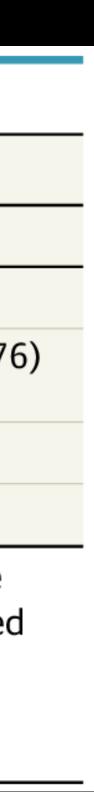
		Time horizons							
Model performance	Hospitalization	24 h	12 h	8 h	4 h				
Outcome incidence, %	6.6	0.43	0.29	0.22	0.14				
Area under the receiver operating 0.63 (0.62-0.64) characteristic curve (95% CI)		0.72 (0.72-0.72)	0.73 (0.73-0.74)	0.74 (0.74-0.75)	0.76 (0.75-0.76				
Positive predictive value (ESM score ≥6), %	12	2.4	1.7	1.4	0.92				
No. needed to evaluate (ESM score ≥6) <sup>a</sup> 8		42	59	73	109				
Abbreviation: ESM. Epic Sepsis Model. only the first time the ESM score is 6 or higher. For each time horizon, the									

ADDIEVIALION: ESIVI, EPIC SEPSIS MOUEI.

<sup>a</sup> The number needed to evaluate makes different assumptions at the hospitalization and time horizon levels. At the hospitalization level, the number needed to evaluate assumes that each patient would be evaluated

## Table 2

only the first time the ESW score is 6 or higher. For each time horizon, the number needed to evaluate assumes that each patient would be evaluated every time the ESM score is 6 or higher.



	Sepsis	No Sepsis	
<b>ESM ≥ 6</b>	843	5,948	6,791
<b>ESM &lt; 6</b>	1,709	29,955	31,664
	2,552	35,903	38,445

Wong et al. 2021

	Sepsis	No Sepsis	
<b>ESM ≥ 6</b>	843	5,948	6,791
<b>ESM &lt; 6</b>	1,709	29,955	31,664
	2,552	35,903	38,445

Wong et al. 2021

 $PPV = \frac{TP}{TP + FP} = \frac{843}{6791} \approx 12\%$ 

	Sepsis	No Sepsis	
<b>ESM ≥ 6</b>	843	5,948	6,791
<b>ESM &lt; 6</b>	1,709	29,955	31,664
	2,552	35,903	38,445

Wong et al. 2021

$$\frac{1}{V} = \frac{6791}{843} \approx 8$$

	Sepsis (No Abx)		
<b>ESM ≥ 6</b>	183	660	843
<b>ESM &lt; 6</b>	679	1,030	1,709
	862	1,690	2,552

Wong et al. 2021

#### Seps (No Ak $ESM \ge 6$ **ESM < 6** 6

183  $P(Useful \cap Correct)$  $\approx 22\%$ P(Useful | Correct) = -843 P(Correct)

Wong et al. 2021

	Sepsis (Abx)	sis ox)
843	660	83
1,709	1,030	79
2,552	1,690	62

### Why such a big difference between expected & observed performance?

#### Subtle choice of outcome definition

Development: ICD-9 code indicating diagnosis of sepsis Our outcome: Health catalyst operational sepsis outcome

Billing lags behind actual clinical care

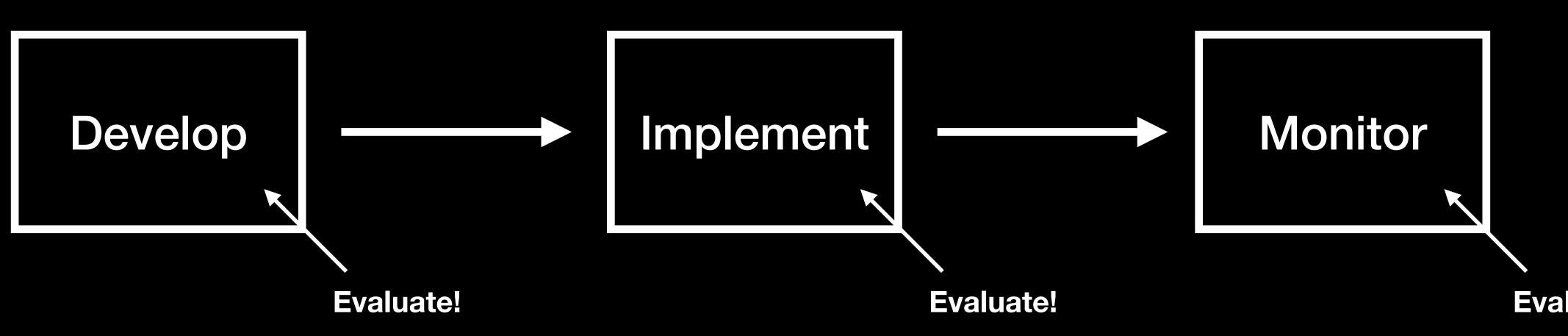
#### makes a big difference

#### Sensitivity Analysis



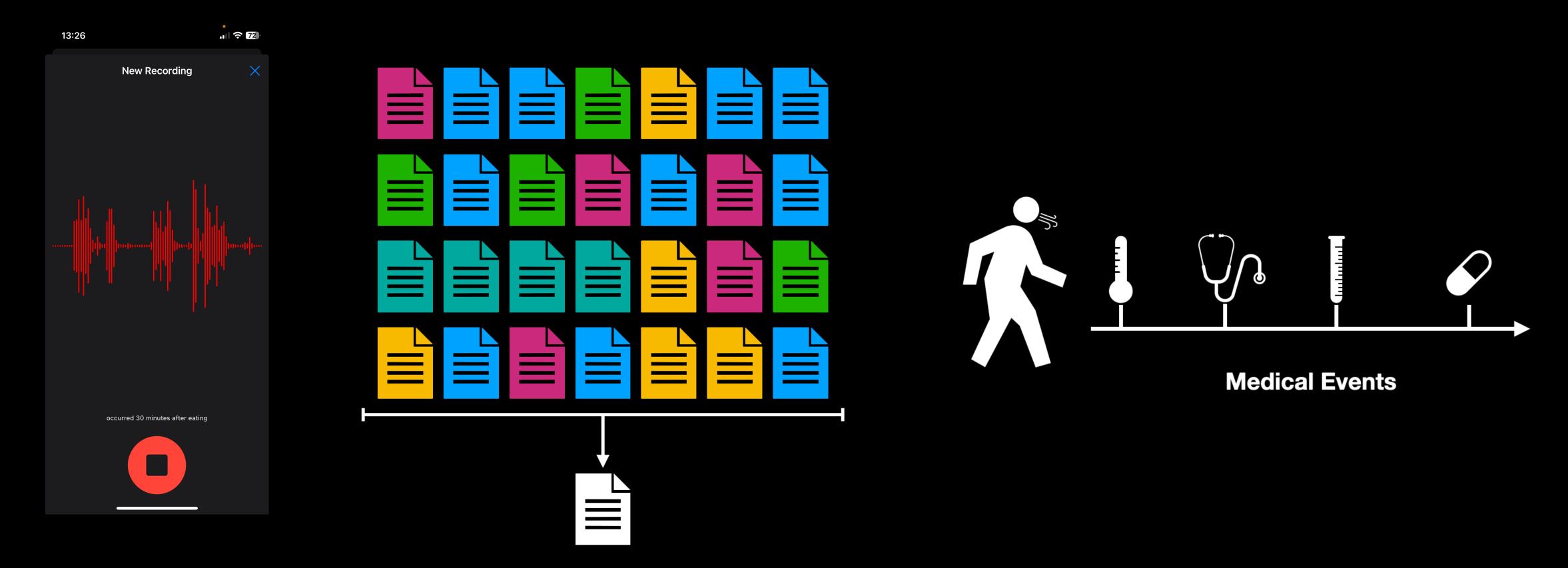
When ESM scores up to 3 hours after the onset of sepsis were included, the hospitalization-level AUC improved to 0.80 (95% CI, 0.79-0.81).

#### All models are wrong, but some are useful...





#### Generative AI Tools Coming Down the Pike



#### Al Scribe

#### AI Chart Summarization



Medical Foundation Models



### Al Scribes

Goals:

Reduce burden of note creation

Facilitate more face-to-face time

Technology:

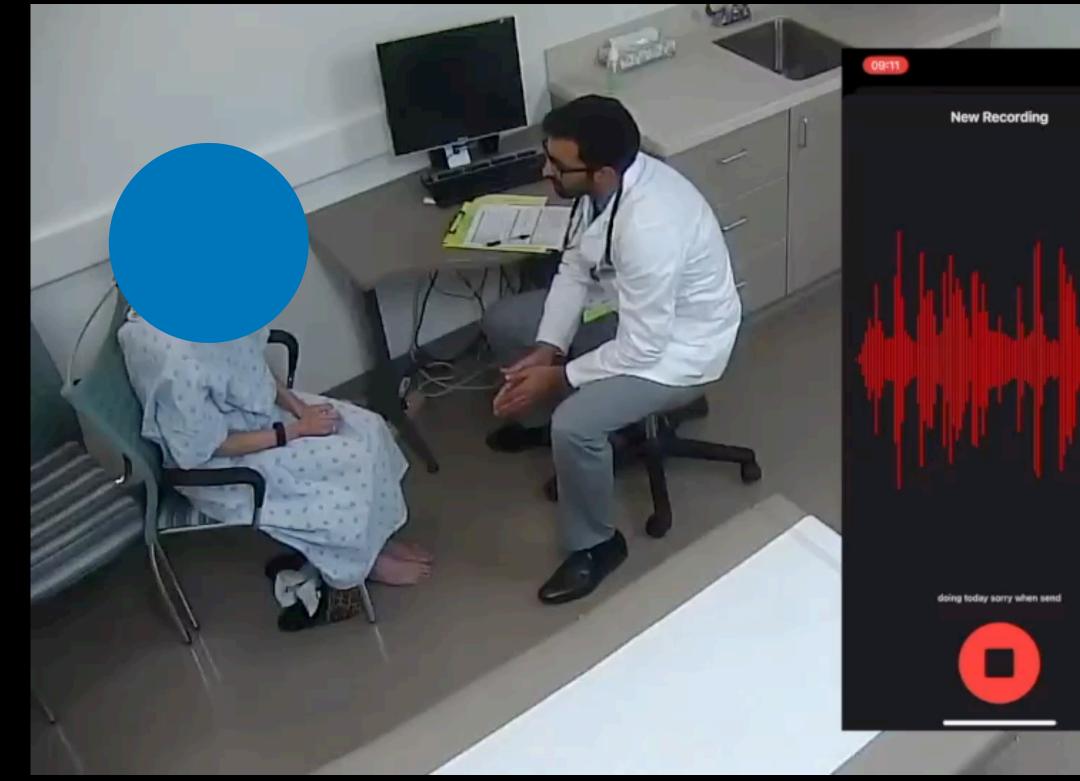
App records encounter

Recording transcribed & converted to a note via LLM

Landscape:

2y ago very hard to build

Now extremely easy to build - hard to validate





### ASCIDES

Goals:

Reduce burden of note creation

Facilitate more face-to-face time

Technology:

App records encounter

Recording transcribed & converted to a note via LLM

Landscape:

2y ago very hard to build

Now extremely easy to build - hard to validate



#### 11:55 Recording Details 🗸 **K** Back Apr 3, 2024 at 09:21

Transcript

Summary

#### \*\*SUBJECTIVE\*\*

00:00

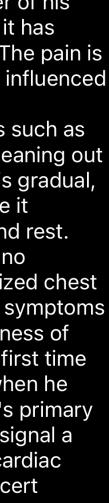
(15)

The patient is a male who presents today with complaints of unusual chest pain. He reports inappropriate discomfort and a vague constricted feeling in the direct center of his chest for the past month, noting that it has occurred approximately three times. The pain is described as a tightness, not notably influenced by any specific factor, but has been experienced during physical activities such as moving boxes in the basement and cleaning out the garage. The onset of discomfort is gradual, persisting for about 10 minutes before it subsides, primarily upon relaxation and rest. Medications have apparently yielded no significant relief. Apart from the localized chest pain, patient reported accompanying symptoms like light sweating, nausea, and shortness of breath that were experienced for the first time yesterday during grocery shopping when he had to carry heavy bags. The patient's primary concern is whether these symptoms signal a serious condition, given his father's cardiac history. He also expressed his disconcert

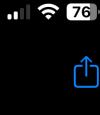
1.93 MB



-10:51







TPMG studied an AI scribe

From unnamed vendor

Accessible to a wide range of physicians

As of publication time

3k physicians, 303k encounters

Studied

PJ time

Time in notes

Note quality

TPMG studied an AI scribe

From unnamed vendor

Accessible to a wide range of physicians

As of publication time

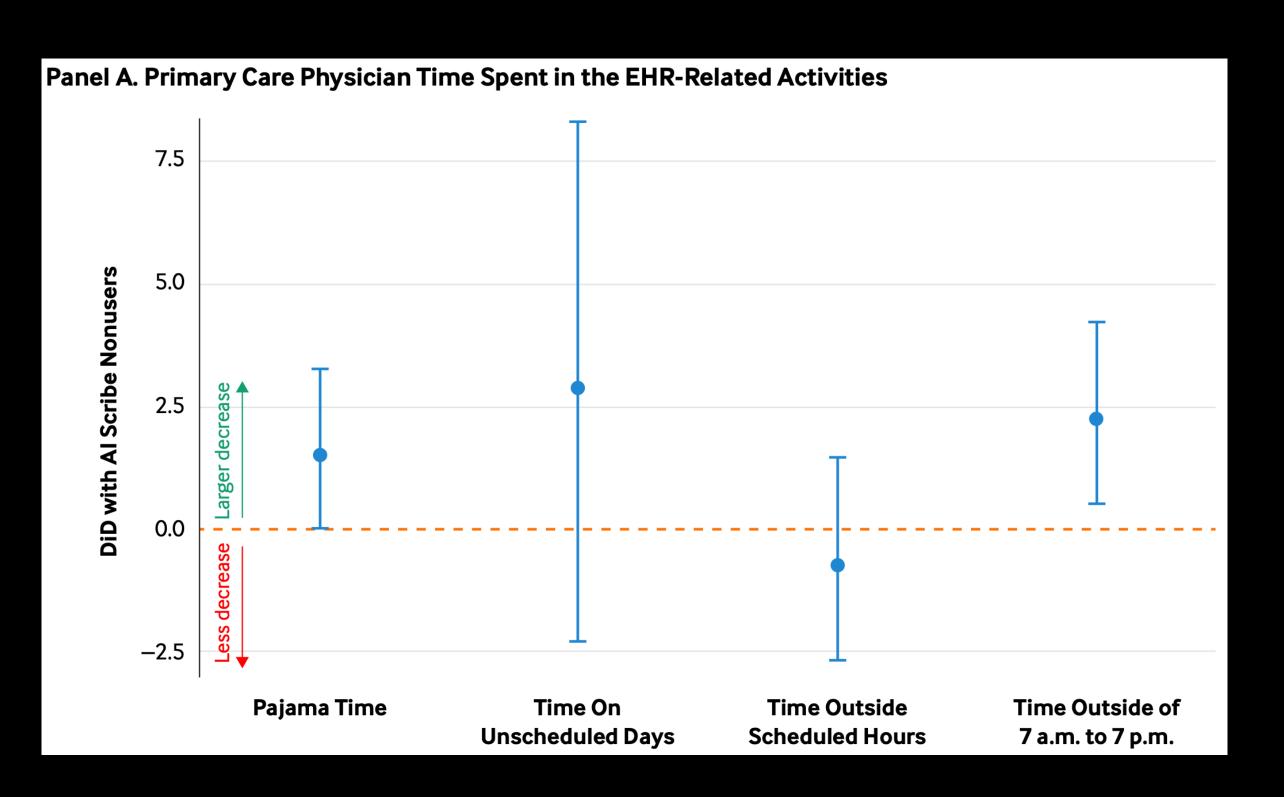
3k physicians, 303k encounters

Studied

PJ time ↓

Time in notes

Note quality



TPMG studied an AI scribe

From unnamed vendor

Accessible to a wide range of physicians

As of publication time

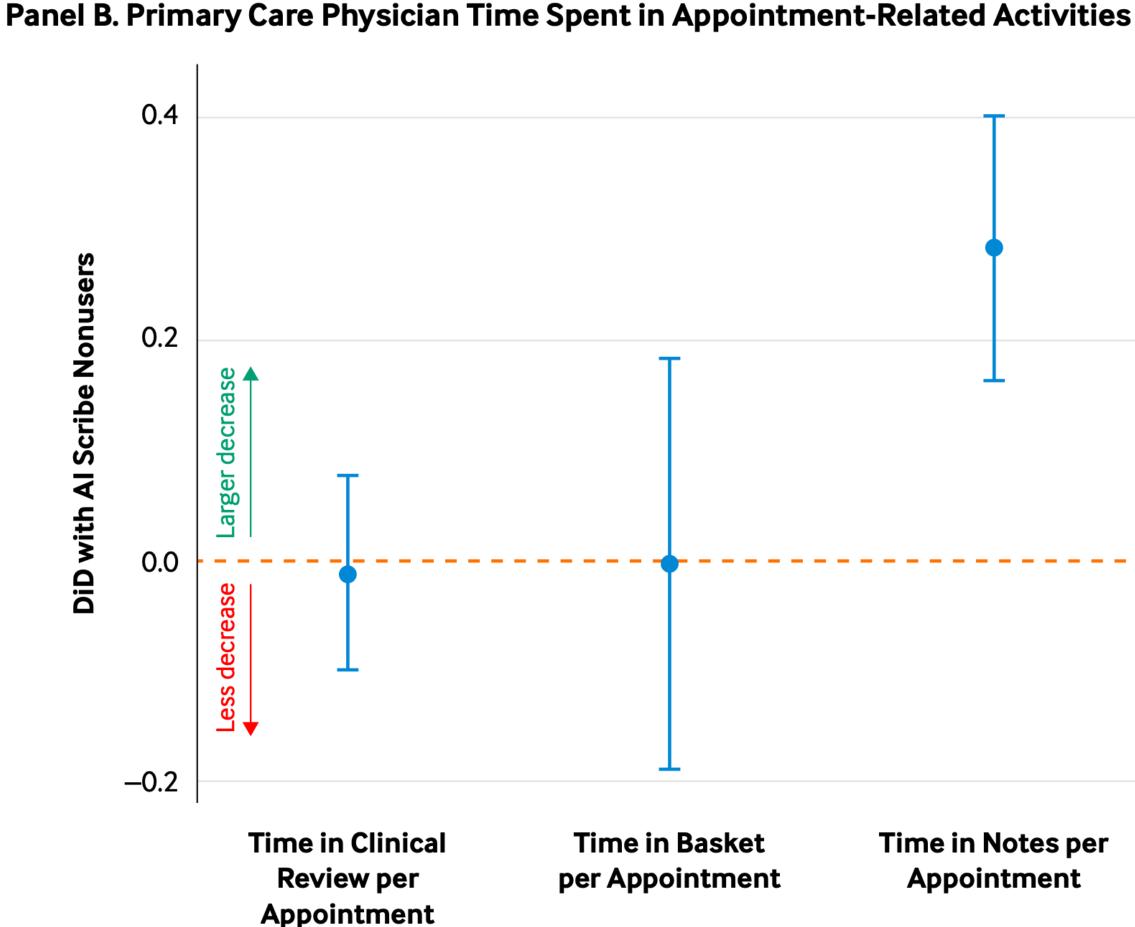
3k physicians, 303k encounters

Studied

PJ time ↓

Time in notes  $\downarrow$ 

Note quality





TPMG studied an AI scribe

From unnamed vendor

Accessible to a wide range of physicians

As of publication time

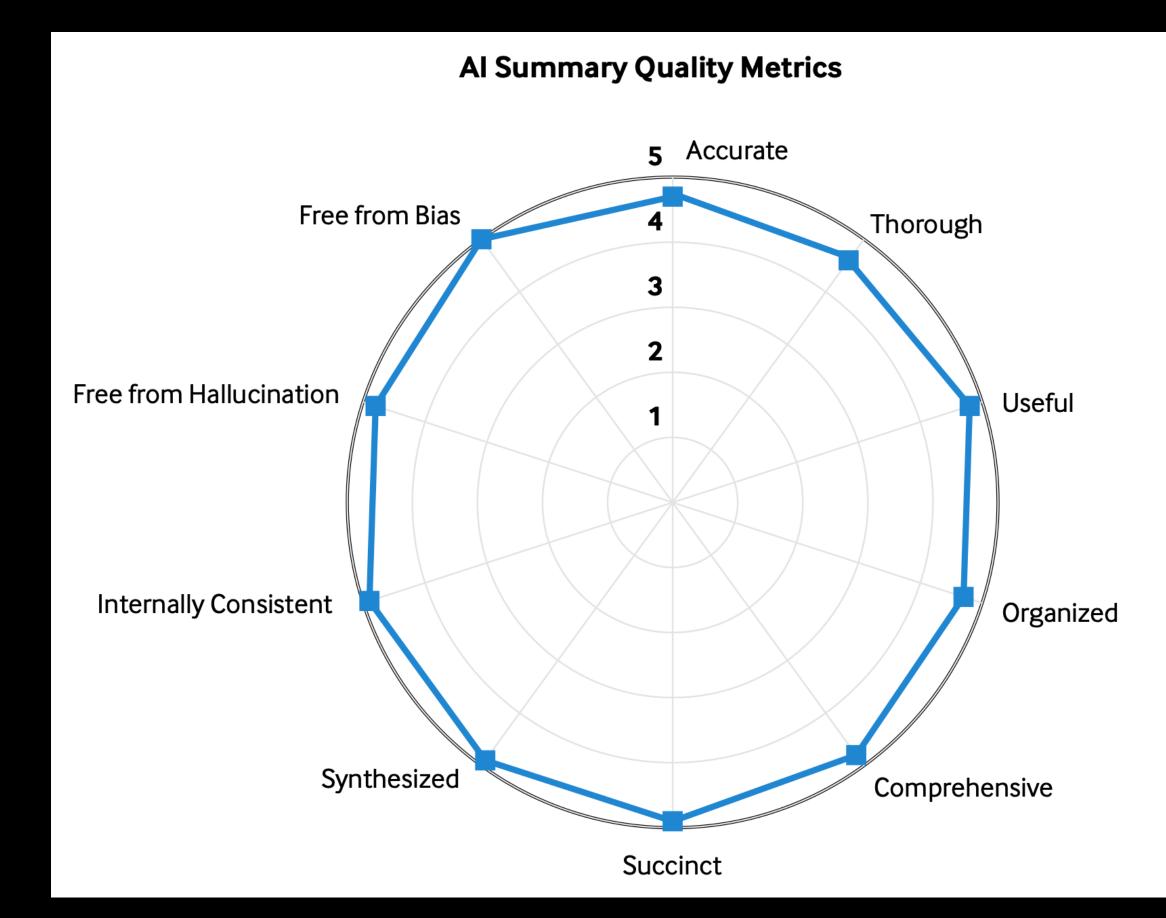
3k physicians, 303k encounters

Studied

PJ time ↓

Time in notes  $\downarrow$ 

Note quality ~



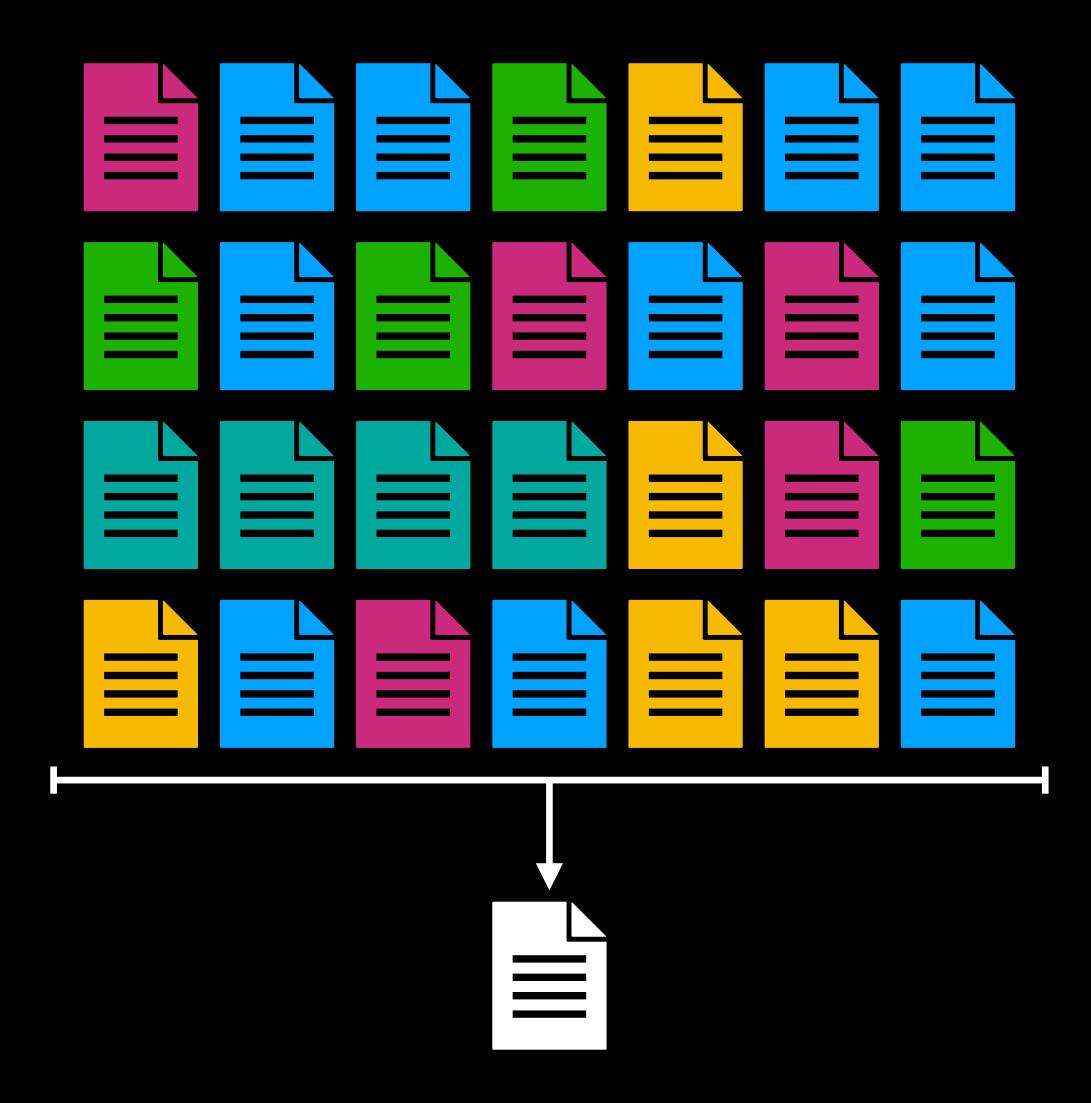


## Al Chart Summarization

Goals:

Reduce burden of chart review Help highlight relevant info Reduce irrelevant info Technology: LLM operating over all available notes Some discussion of clinician-in-the-loop Landscape:

Easy to build very hard to validate



## **Example Chart Summarization**

	r 20 e Vi	24, Mo sit	nth 1	.: Pri	mary			Year Car	2024	, Mon	th 10	):			
Su	Year Up P	2024, rimary	Mont Care	h 4: Visi	Follow- t			Sub	Year Telep Prima	hone	Εηςοι				
Ch fc Ji pr Je or di ge hy Li Al Fe Sc al wi	Sub Chi for try med His son rea dil diz diz diz diz diz diz diz diz diz son con con con con con con con con con c	Year 20 Endocri Su Yea Hos Ch Not ca Ch Not ca Ch Adm Day Ch Day Chi Day Chi Sut Sut Sut Sut Sut Sut Sut Sut Sut Sut	224, inolo r 202 pital Year Hosp Prog Sub Chi fee Int sig fol Den sho sym dia hea pro Obj Vit 98. Phy Gen adm	Month gist 24, Mc lizati 2024 oitali ress Year Hospi Progr Card Subj Chie feel over Inte cath reve arte (PCI perf reli proc be n of b inse Obje Vita 98.6 Phys		2 Daily 3 - Post- nth 9: on - Discharge sion: [Month 9 024] arge: [Month 9 024] gnosis: Elevation Myoc rformed: terization wit rvention (PCI) two coronary a se: -year-old male betes, hyperte a, was admitted th NSTEMI. Car on revealed two els, which wer d with stents. urse was uncom	cardial Infa cardial Infa ch percutane and stent arteries. e with a his ension, and ed with ches diac vo partially re successfu His post- plicated, w	Chi for up. ste Int to exp sid sta and lev his lon Rev sho diz r o f f arcti eous story st pa	Subje Chief Feler feeli Ista Inter well- regin menti prima weeks synco short Revie objec Vital Asses	Year Neph Subj Chie doct my k Hist no s suck legs epis sind medi reg	2025 rolog Year Hosp Note Date 2025 Sub: Chie tire swo His year dial hype 3 Chie tire swo His year dial hype 3 Chie tire swo His year dial hype 3 Chie tire swo His year dial hype 3 Chie tire swo His year dial hype 3 Chie tire swo His year dial hype 3 Chie tire swo His year dial hype 3 Chie tire swo His year dial hype 3 Chie tire swo His year dial hype 3 Chie tire swo His year dial hype 3 Chie tire to to to to to to to to to to to to to	202 ital Yea Hos Pro Dat Day Sut Chi the Int inf inf cor ove Obj	25, Mo lizati r 2029 pital: gress Year Hosp: Summa Date 2025 Date 2025 Admi Acute Hear Frace Hype Hosp: Mr. I sugge bila breat	<pre>ultation onth 6: on - Admission 5, Month 6: ization - Dail Note (Dav 3) 2025, Month 6 italization - Year 2025, Mo Care Follow- Subjective: Chief Complation for a routing to see how I a few months Interval Histimprovement He's been add regimen, dietintake recomming cardiac rehal with regular ongoing conce and managing conditions. Review of System attributes to extremities, Positive for attributes to</pre>	-y : Discharge onth 9: Pri unt: "I'm h check-up 'm doing af ago." tory: Mr. D in his symp hering to h tary restri mendations. bilitation physical a ern about h his multip stems: Nega nificant sw or episode mild fatig o his incre

#### Summarization Example

ChatGPT

J. Doe, a 58-year-old male with a complex medical history, presented to the ED on 10/02/2025 with severe right upper quadrant (RUQ) pain that started 30 minutes after dinner. His medical background includes:

- Type 2 Diabetes Mellitus: Managed with a regimen that includes Metformin and a GLP-1 receptor agonist. His glycemic control has been generally good, with a recent HbA1c of 6.9% [11].
- Hypertension: Well-controlled on a regimen including Lisinopril and Metoprolol succinate. His blood pressure was 132/82 mmHg at the last primary care follow-up [77]
- Hyperlipidemia: Managed with Atorvastatin 80 mg nightly, achieving LDL levels within the target range [1].
- Coronary Artery Disease: History of NSTEMI managed with percutaneous coronary intervention (PCI) and stent placement, currently on dual antiplatelet therapy with Aspirin and Clopidogrel [17].
- Chronic Kidney Disease (CKD) Stage 3: Diagnosed around March 2025, likely secondary to his long-standing hypertension and diabetes. His management includes close monitoring of renal function and adjustments to medications to minimize nephrotoxicity. Recent serum creatinine stabilized at 1.7 mg/dL, with an eGFR of approximately 46 mL/min/1.73m<sup>2</sup> [1].
- Heart Failure with Preserved Ejection Fraction (HFpEF): Diagnosed in June 2025 following an admission for symptoms suggestive of volume overload, including shortness of breath and bilateral ankle swelling. His heart failure is managed with Furosemide, Lisinopril, and Metoprolol succinate, emphasizing sodium and fluid restriction [**''**] [**''**].
- Orthostatic Hypotension: Episodes noted, particularly related to antihypertensive therapy adjustments [1].

In recent months, Mr. Doe has generally reported improvement in his symptoms, attributing mild fatigue to increased activity levels. He's been adhering to his medication regimen, dietary restrictions, and fluid intake recommendations, completing a cardiac rehabilitation program, and maintaining regular physical activity [17].

Given his RUQ pain, considerations might include gallbladder issues (e.g., cholelithiasis, cholecystitis) given the postprandial onset of symptoms, as well as possible hepatic or pancreatic involvement. His history of CKD and HFpEF, coupled

er my hospital stay

reports general oms since discharge. s medication tions, and fluid He completed a rogram and continue tivity. He expresse s kidney function e chronic

ive for chest pain, lling of the of dizziness. e, which he sed activity level.

R 72 bom. Weigh



## **Clinical Foundation Models**

Goals:

Build general models that can help us answer many different clinical questions

Operate over physiologic values & medical events instead of words

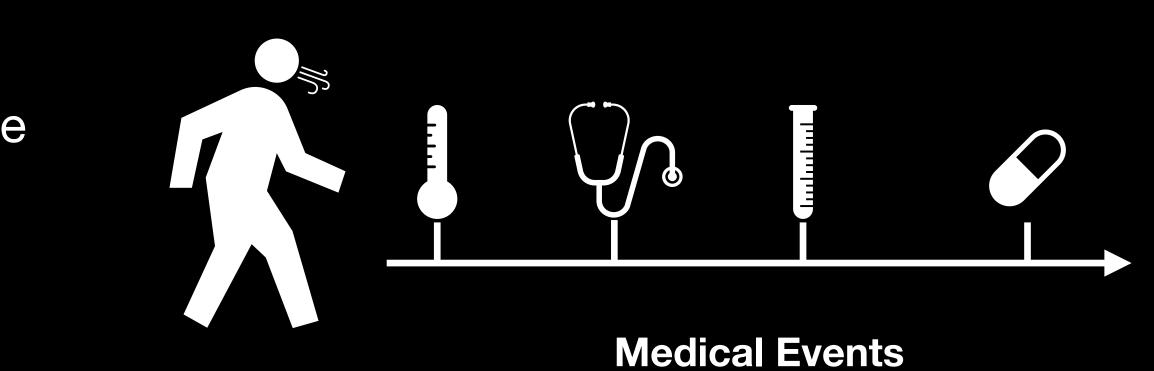
Technology:

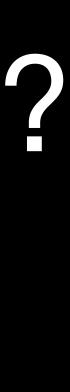
Transformer operating over all available EMR data

Landscape:

Hard to build hard to validate

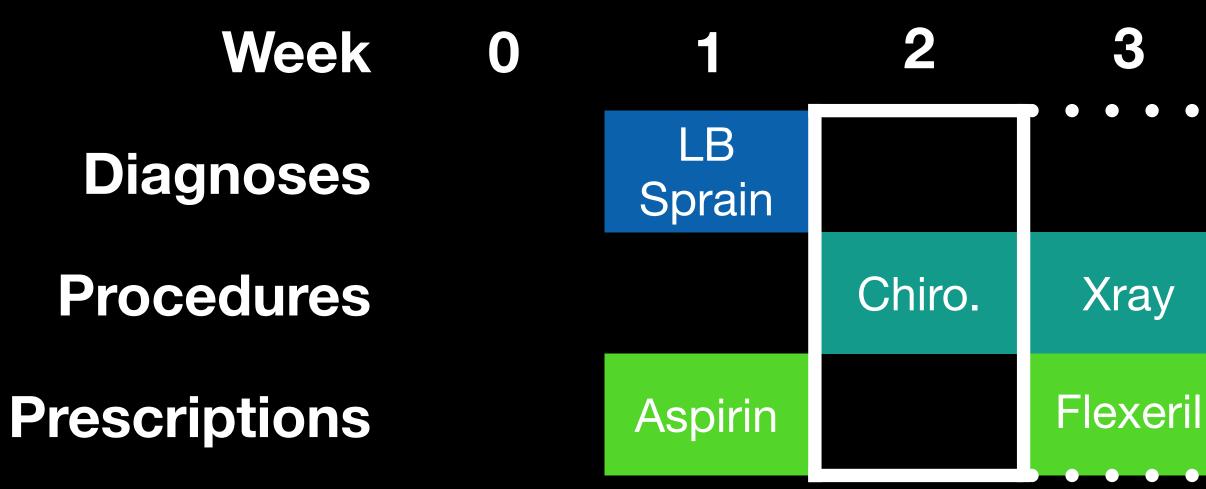
Natural Langauge The quick brown fox jumps over ... ?







#### Similar to Recovery Trajectory Generation



Doesn't involve repurposing another AI model. Need to make special medical AI models.

**Resource intensive** 

Computationally expensive

Massive data needs

Specialized engineering skills

Need to keep in mind

Privacy

Bias

Interprebility

Maintenance

### Takeaways

Generative AI is special case of AI

Having a general understanding of AI aids in understanding generative AI

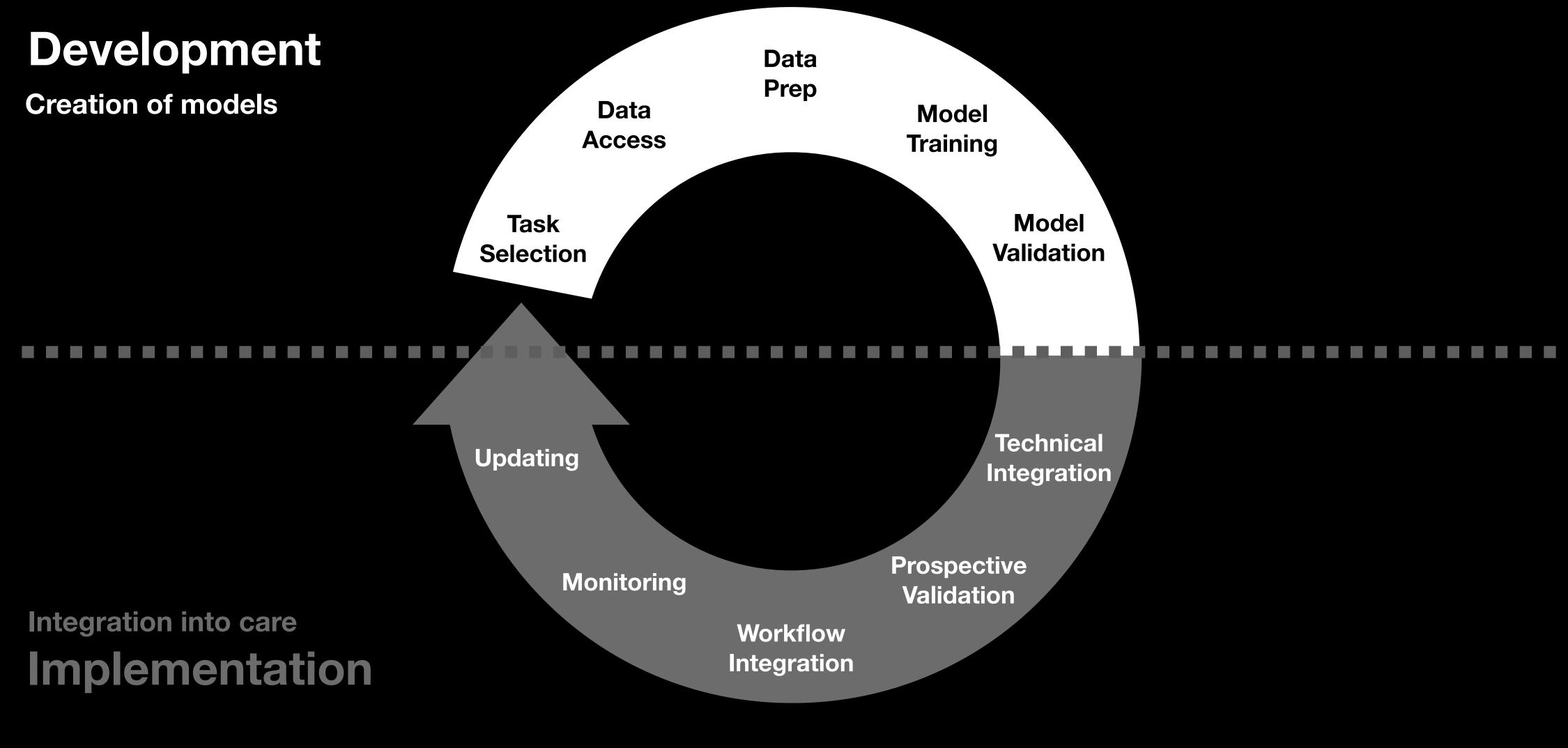
Models can be used in both a generative and predictive sense

Evaluation is critical in Al

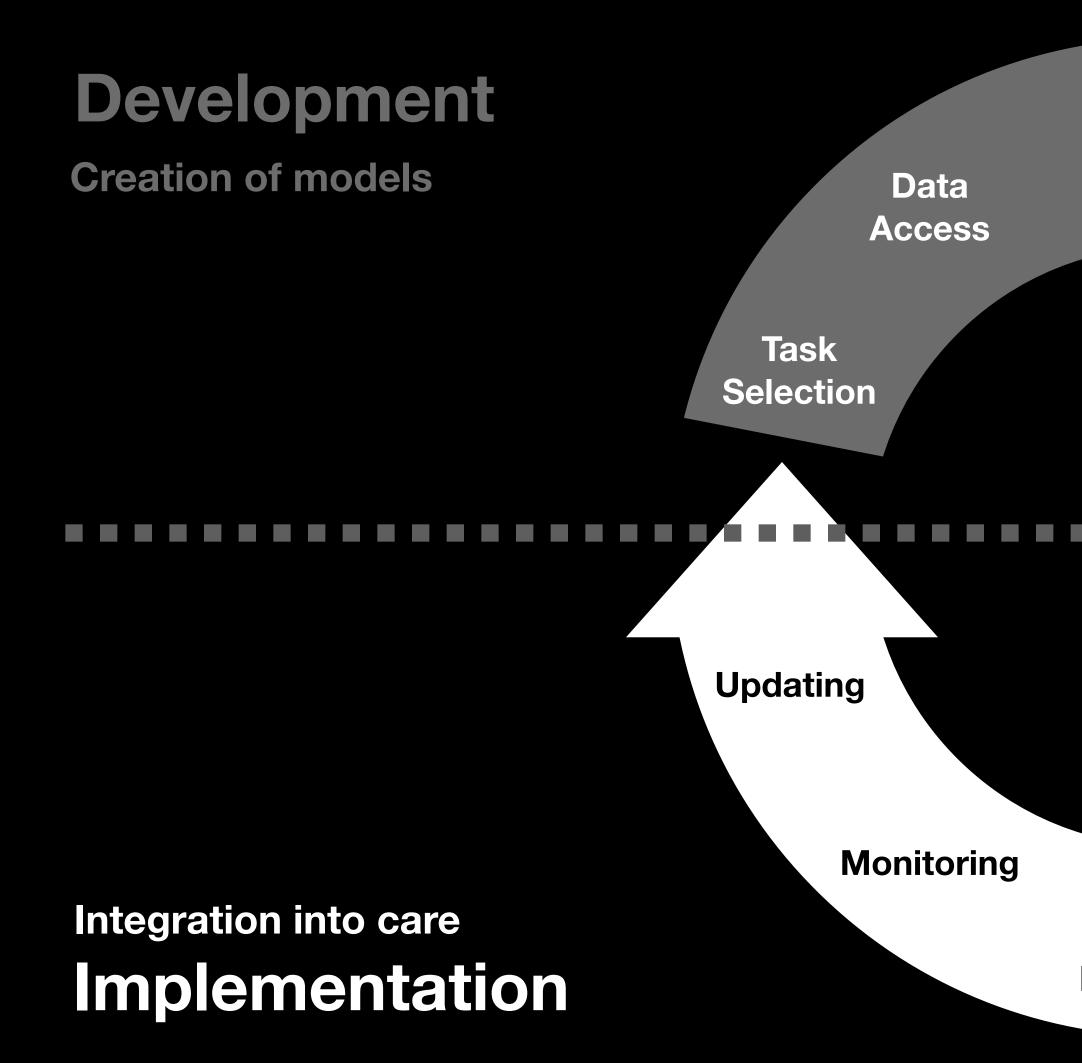
Generative AI is harder to evaluate because the larger amount of use cases

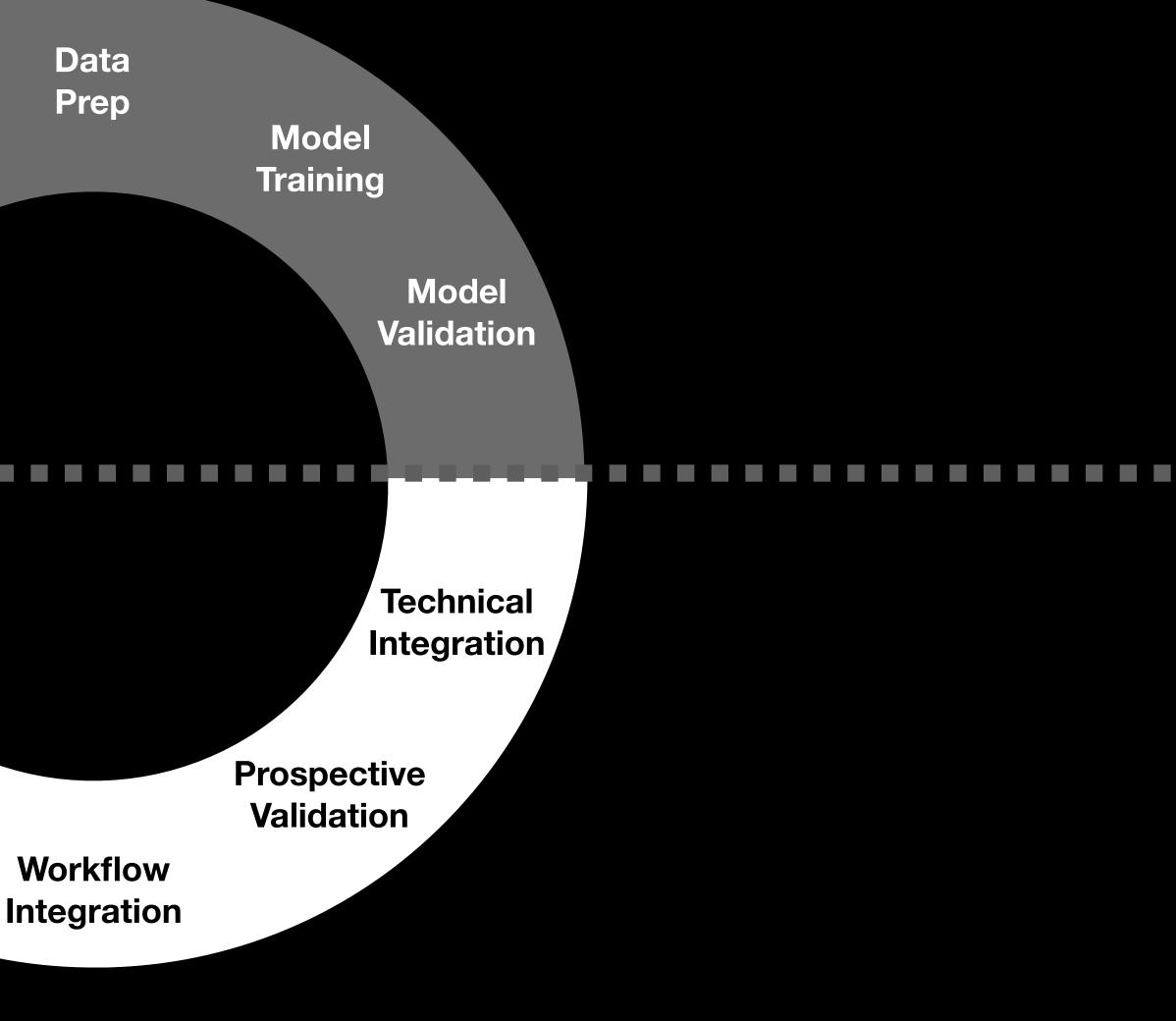
Population biases may be harder to detect

#### Most of these tools are still in development



#### Physicians need to drive the implementation





#### Questions? **Comments? Concerns? Discussion.**

Erkin Ötleş Twitter: @eotles eotles.com eotles@umich.edu





