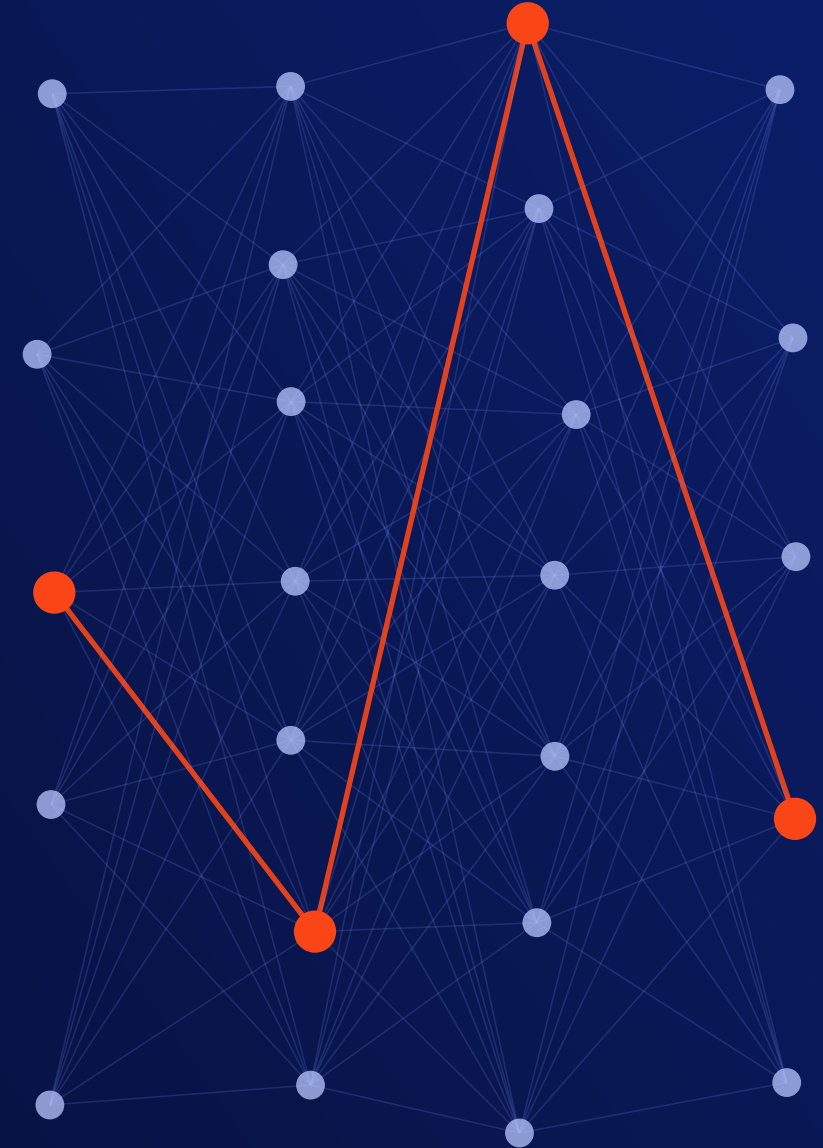

LECTURE 8-1

Foundation Models for EHR and Time Series

Patient timelines and physiological signals as tokens



Parisa Rashidi · Herbert Wertheim College of Engineering, University of Florida



Where We Are

Weeks 6 to 8 apply foundation models to biomedical data. Today: the patient record.



Why the Patient Record Is Different

Not a sentence and not an image: codes, numbers, and time, recorded irregularly.

Codes

diagnoses, drugs, procedures from vocabularies with tens of thousands of entries

Numbers

labs and vitals, each with its own units and normal range

Time

events at irregular intervals; gaps of minutes or months

Missingness

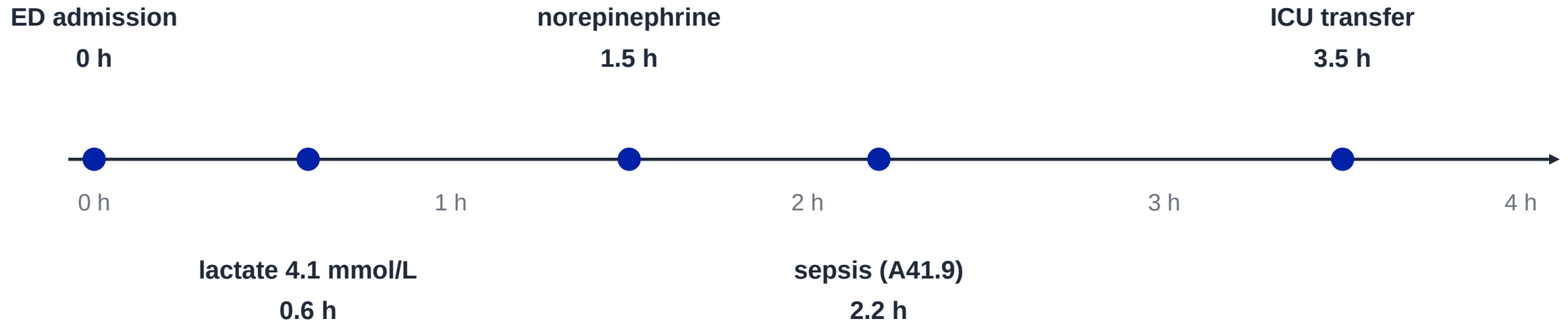
a test that was not ordered says something too



A Patient Timeline as Tokens

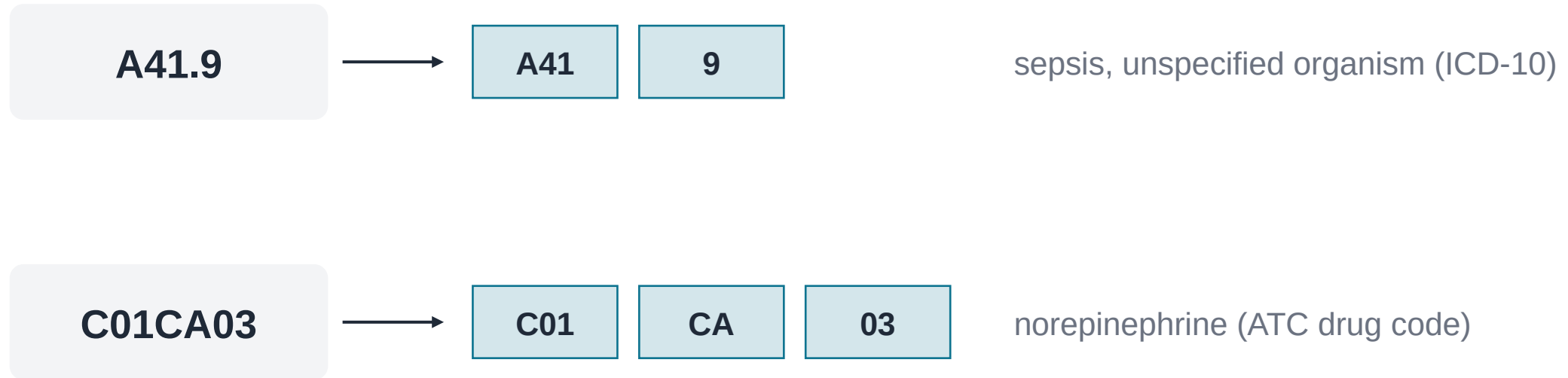
A Record Is a Timeline

One synthetic patient: five events in three and a half hours.



Codes Become Tokens

ETHOS splits each code into its parts, from broad group to specific item.



Broad group first, specific item last: [A41] is shared by every A41 diagnosis.

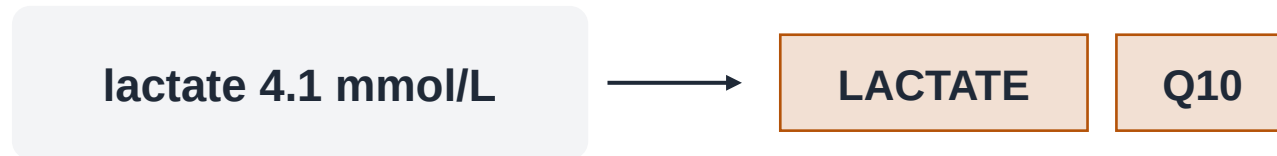
Numbers Become Quantile Tokens

Each lab is cut into 10 bins with equal counts in the training data.

lactate values in the training data, cut into 10 equal-count bins



cutoff ≈ 3.6 (illustrative)



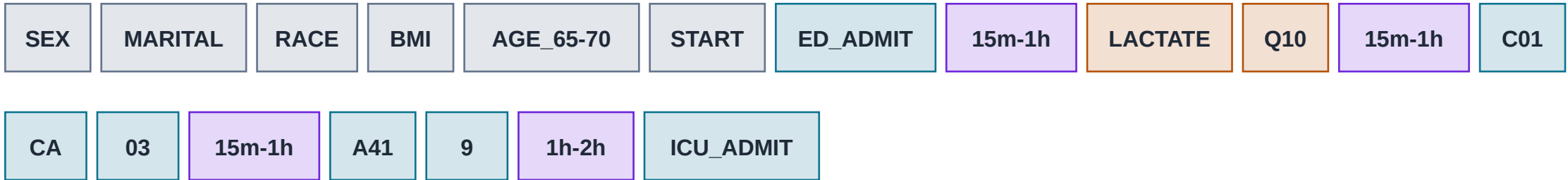
Time Gaps Become Interval Tokens

ETHOS uses 13 interval tokens and adds none for gaps under 5 minutes.

5m-15m	15m-1h	1h-2h	2h-6h	6h-12h	...
ED admission → lactate 4.1 mmol/L	0.6 h	15m-1h			
lactate 4.1 mmol/L → norepinephrine	0.9 h	15m-1h			
norepinephrine → sepsis (A41.9)	0.7 h	15m-1h			
sepsis (A41.9) → ICU transfer	1.3 h	1h-2h			

The Whole Record as One Sequence

Six static tokens, then the events in time order: 19 tokens.



19 tokens: 6 static + 13 for five events and four gaps

Train a Decoder on Timelines

The same next-token loss as Weeks 3 to 5, on hundreds of thousands of records.

ETHOS (Renc et al., 2024)

data: MIMIC-IV, one academic center
241,015 patients for training and validation
26,758 patients held out for testing
decoder-only, next-token loss

Context

sliding window of 2,048 tokens
oldest tokens drop off as the
timeline grows
dense records lose early history

Animation: A Patient Timeline as Tokens

Animation clip: week08-clip1.mp4

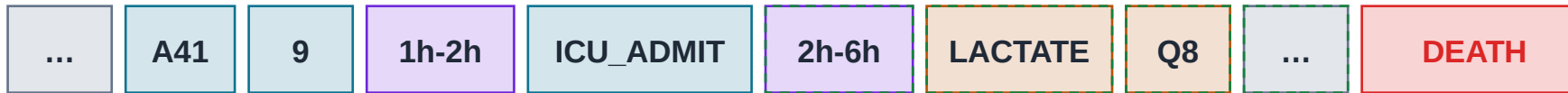
Silent version embedded (click to play). The narrated version is in the course repository.



Zero-Shot Prediction by Simulating the Future

Ask a Question Without Training

Will this patient die in the ICU? Continue the timeline until it ends.



the record so far

sampled by the model (illustrative)

Stop at a terminal token

DISCHARGE or DEATH ends the simulated future;
the interval tokens along the way give the simulated length of stay.

Twenty Futures

Sample 20 continuations; the share that ends in death is the predicted risk.



7 of 20
end in death

risk = 0.35

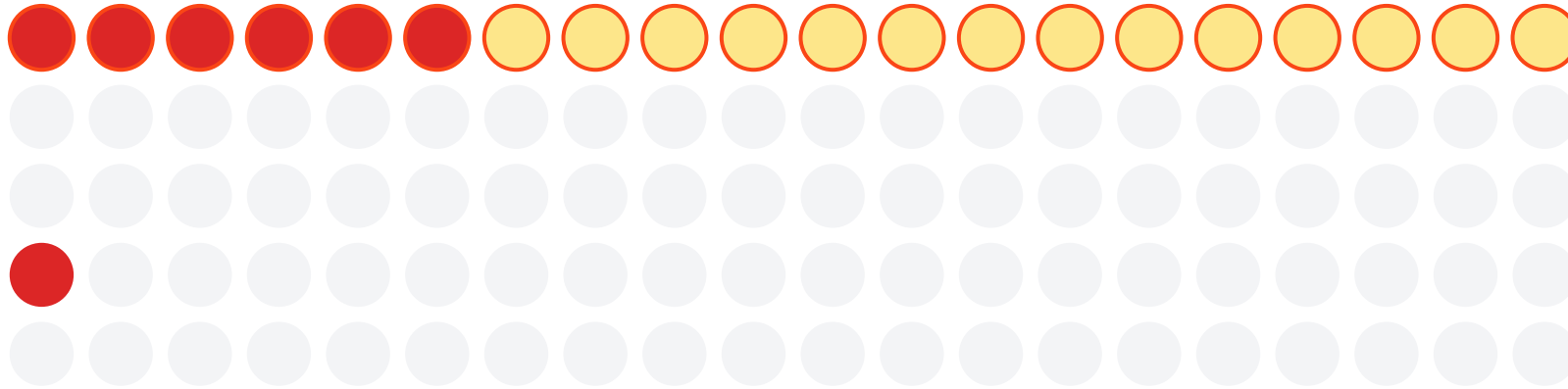
What ETHOS Reports

Strong discrimination across tasks, from one model and no task-specific training.

ICU mortality	AUC 0.927	7.0% prevalence, 7,483 ICU stays
In-hospital mortality	AUC 0.921	2.0% prevalence
ICU readmission	AUC 0.807	
Diagnosis-related group	84.8% top-1	771 classes

High AUC, Many False Alarms

When few patients have the outcome, most alarms can still be false.



red: died (7) · orange ring: flagged (20)

20 flagged

6 of 7 deaths caught
sensitivity 86%

14 false alarms
PPV 30%

Animation: Zero-Shot Prediction by Simulating 20 Futures

Animation clip: week08-clip2.mp4

Silent version embedded (click to play). The narrated version is in the course repository.

How to Read EHR Foundation Model Results

Ask where the comparisons, the data, and the tasks come from.

Comparisons

baselines reported by other studies, with different cohorts

Data

one academic center (MIMIC-IV); no external site

Context

dense records lose their early history

Tasks

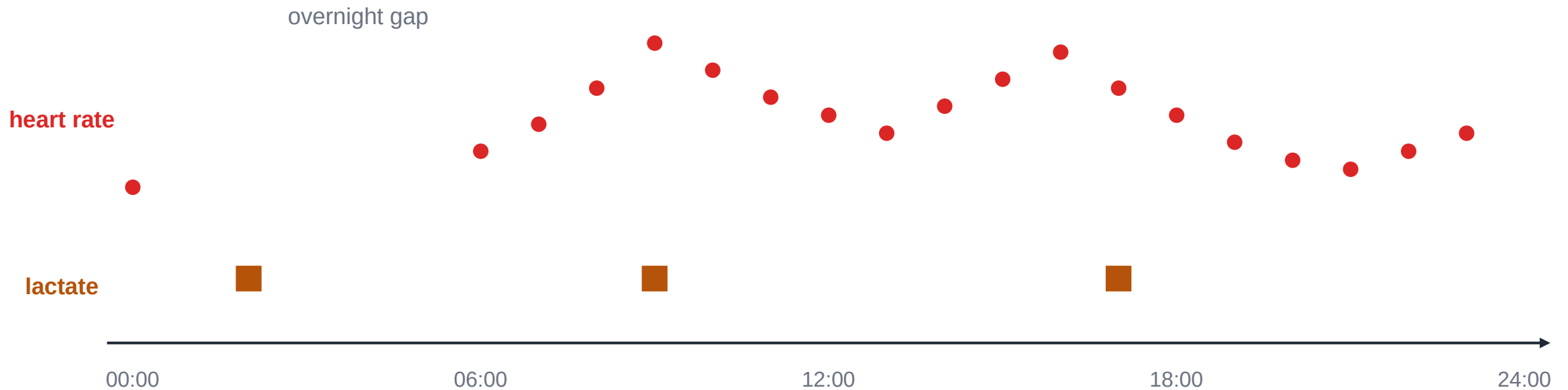
Wornow et al.: benchmarks say little about value to a health system



Irregular Sampling and Physiological Signals

Measurements Arrive Irregularly

Heart rate about hourly with an overnight gap; lactate only three times a day.



Note the 02:00 lactate: drawn during the gap in heart-rate charting.

Option 1: Grid, Fill, and Mask

Put values on an hourly grid, fill gaps with the last value, and record a mask.

	00:00	01:00	02:00	03:00	04:00	05:00	06:00
heart rate	88	88	88	88	88	88	92
mask	1	0	0	0	0	0	1
lactate	—	—	4.1	4.1	4.1	4.1	4.1
mask	0	0	1	0	0	0	0

bold: measured · grey: filled in · mask 1 = measured

Missingness Carries Information

Nobody draws a lactate at 2 a.m. without a reason.

Signal

an off-hours lactate
signals clinical concern
the mask lets the model use it

Risk

ordering habits differ between
hospitals and change over time;
a model can learn the habit,
not the patient (Week 12)

Options 2 and 3: Interval Tokens and Time Embeddings

Make time itself an input: as a token, or as waves in real hours.

Option 2: interval tokens

the gap is a token:

[2h-6h]

used by ETHOS

Option 3: time embedding

$t = 2.5$ hours

8-hour wave: $\sin 0.92, \cos -0.38$

24-hour wave: $\sin 0.61, \cos 0.79$

like Week 2 positions, in real hours

Dense Signals as Patches

A 10-second ECG at 500 Hz becomes 10 one-second patches of 500 samples.



10 patches × 500 samples → 10 tokens for an ECG foundation model

Animation: Irregular Sampling and Physiological Signals

Animation clip: week08-clip3.mp4

Silent version embedded (click to play). The narrated version is in the course repository.

Every Clinical Data Type Can Become Tokens

Data	Becomes	What is lost or risked
Codes	hierarchical tokens ([A41] [9])	rare codes get little training signal
Lab values	quantile tokens ([LACTATE] [Q10])	exact value and units
Time	interval tokens or time embeddings	precision within a bin
Missing values	a mask, or simply no event	site-specific ordering habits
Dense signals	patches (10 × 500 samples)	detail finer than a patch

Recap: Patient Records as Tokens

Tokens

Codes in parts, values as quantiles, gaps as interval tokens, static tokens first.

Simulate

20 sampled futures give a risk with no task training; AUC 0.927 for ICU mortality in ETHOS.

Caution

Reported baselines, one site, the context window, and many false alarms at low prevalence.

Time

Masks, interval tokens, time embeddings; dense signals as patches.

Before Thursday

Read

Wornow et al., The shaky foundations of LLMs and foundation models for EHRs (2023) and Renc et al., ETHOS (2024).

Write

Reading responses before 1:00 PM Thursday; project peer reviews by 11:59 PM Thursday.

Explore

Week 8 reference notebook (ungraded): a small ETHOS on synthetic ICU stays, rollouts, masks, time embeddings, ECG patches.

Next week

Retrieval, tool use, and agentic systems.