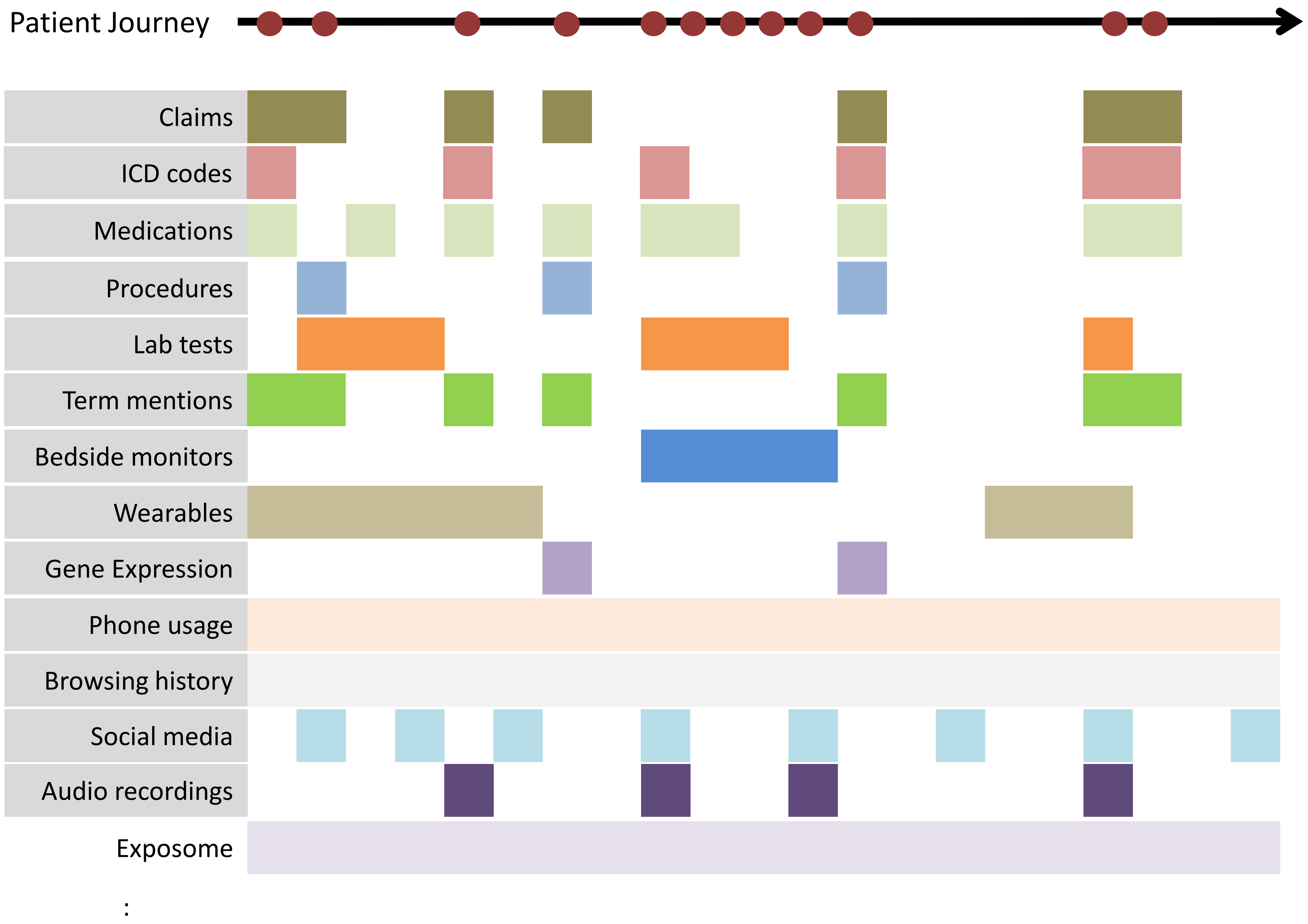
How to Trust, but Verify, in Healthcare

Nigam Shah, MBBS, PhD
nigam@stanford.edu

STANFORD
SCHOOL OF MEDICINE


Patient Journey
Claims
ICD codes
Medications
Procedures
Lab tests
Term mentions
Bedside monitors
Wearables
Gene Expression
Phone usage
Browsing history
Social media
Audio recordings
Exposome
:

Based on:

1. Genetic markers
2. Demographics & SES
3. Prior medical record
4. Wearables (digital biomarkers)
5. Behavioral and social data

Decide who to treat

If (Risk > Th.)

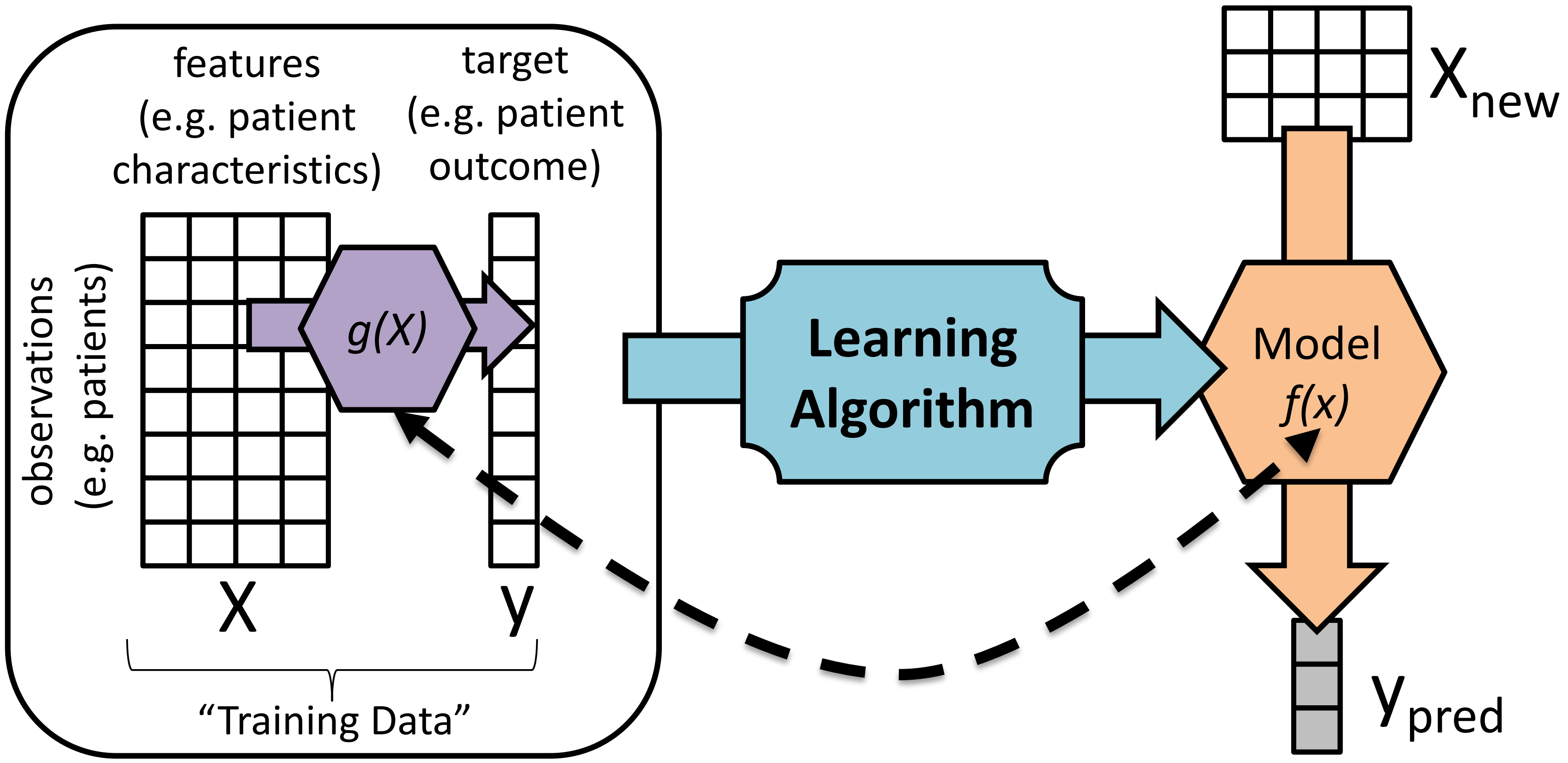
then (do = X)

Decide how to treat

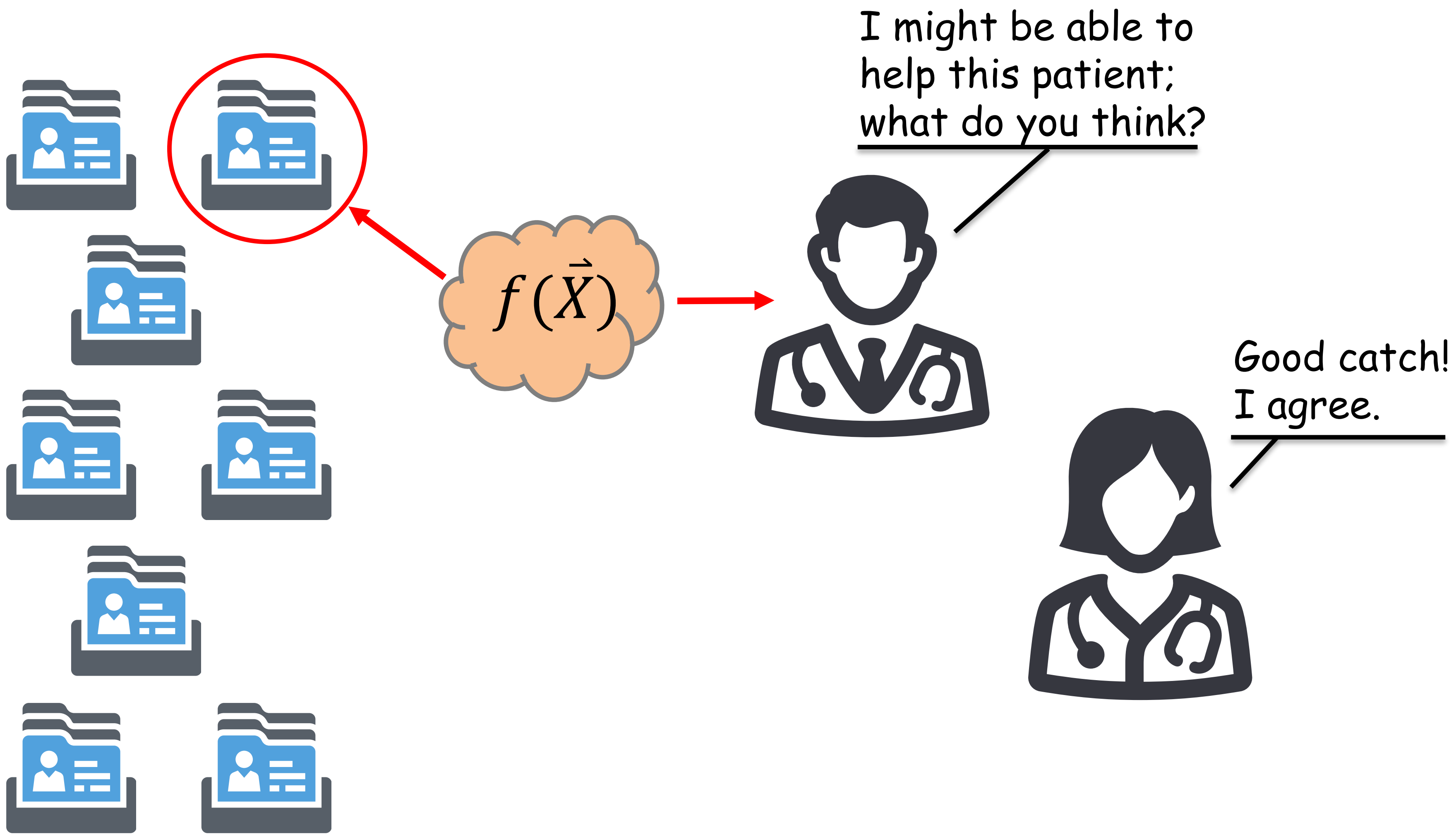
Based on:

1. Biomarker measurements
2. Mechanistic understanding of disease
3. Similar patients’ outcomes
4. What’s covered, and available
5. How much time we have on hand

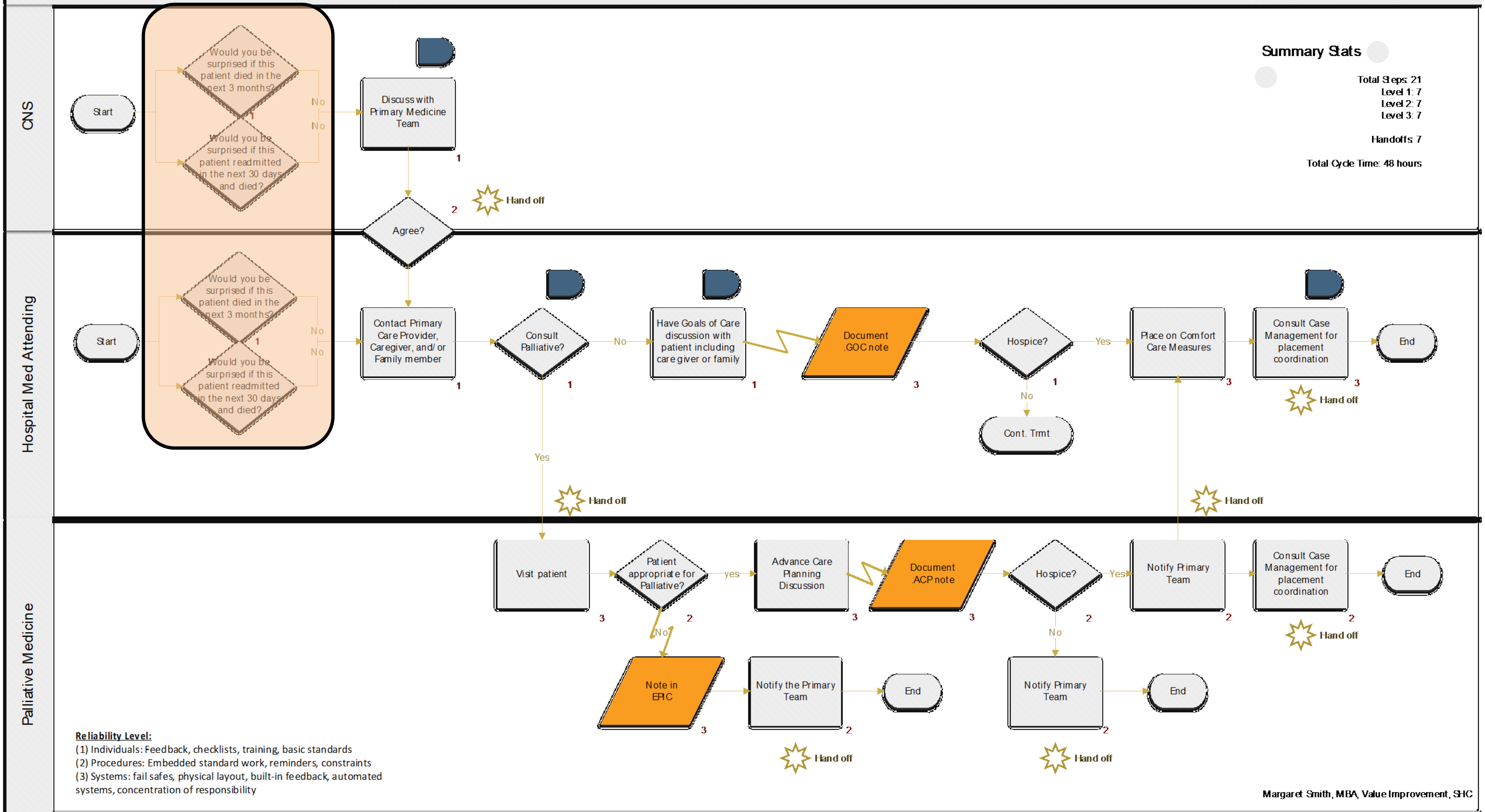
A model


features
(e.g. patient characteristics)
target
(e.g. patient outcome)
observations
(e.g. patients)
g(X)
X
y
"Training Data"
Learning Algorithm
Model
f(x)
X_{new}
y_{pred}

A workflow


$f(\vec{X})$
I might be able to help this patient; what do you think?
Good catch! I agree.

Inpatient Hospital Medicine Initiated Workflow – Goals of Care Workflow



Definitions and Clarifications

- Trustworthiness: of the model, or the workflow around it, or both?

- Trust = proof over time that a thing does what it claims to do. Trust is earned [over time].

- HOW = interpretability
- WHY = explainability

When predicting 24 hr. mortality …

- Interpretability is a poor surrogate for trust
  - Knowing ‘how’ does not help you decide what action to take

- Explainability is a poor surrogate for trust
  - Knowing ‘why’ does not help you decide what action to take

- Knowing that the model’s prediction has helped make good decisions in the past 2 years.

Building trustworthy (and useful!) models

How do we get the best f: X -> Y?

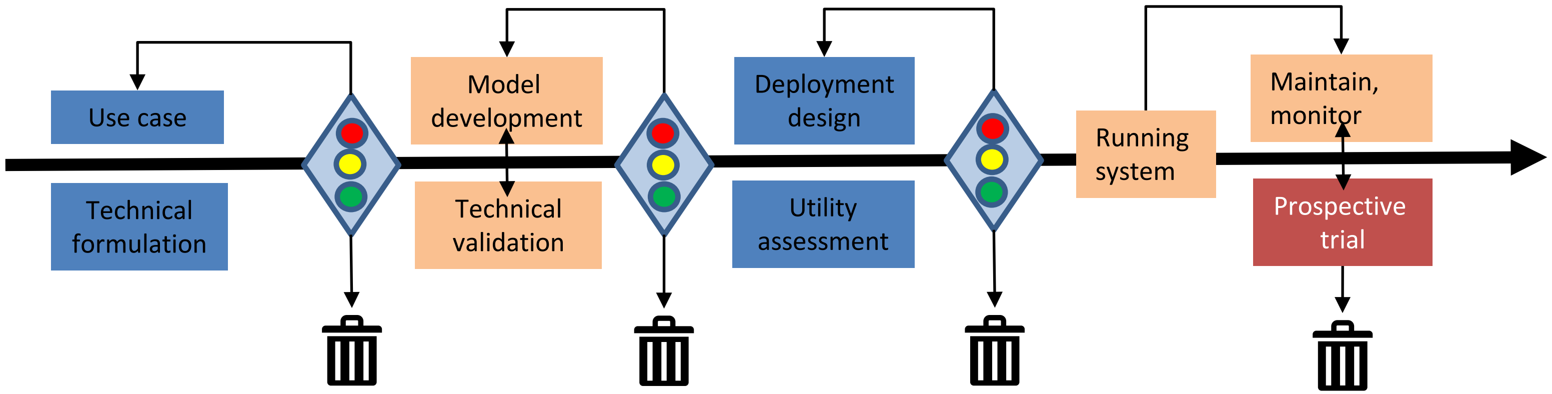
Does representation learning help?
Does multi-task learning help?
Does using textual content help?
How do we train fair models?

Can we use f: X -> Y in the real world?

Can we get the data by 5 am, to make prediction by 6 am?

Running system = model applied to each case + execution of workflow.

- Evaluate the impact of the *running system* on the outcomes we care about
- Maintenance is huge liability – who will carry the pager?
- Monitoring is unexplored



Use case

What clinical outcome(s) are you trying to affect?
Who is the target population?
What action would you take?
Who will take that action?

f: X -> Y subject to…

use an existing equation vs. learn a new equation.

Utility assessment

Given the costs of the actions and its benefit, is there net utility?

Deployment design

Do we increase the efficiency of existing workflows
Do we require entirely new workflows

Acknowledgements

Group Members:

- **Scientists and Instructors:** Ken Jung, Alison Callahan, Jason Fries, Saurabh Gombar, Steven Bagley, Adam Miner
- **Fellows and Visitors**: Ragan Hart, Adrien Coulet
- **Engineers:** Vladimir Polony, Jose Posada
- **BMI Students:** Stephen Pfohl, Sehj Kashyap, Minh Nguyen, Scotty Flemming, Erin Craig
- **MD, EE, and CS Students:** Daisy Ding, Ethan Steinberg, Steve Yadlowsky, Tony Duan

Funding:

- NIH – NLM, NHLBI, NCATS
- Stanford Internal – Dept. of Medicine, Population Health Sciences, Clinical Excellence Research Center, Dean’s office, Stanford Hospital (CEO’s office)
- Fellowships – Med Scholars, Stanford Graduate Fellowship, NSF
- Industry –Janssen R&D, Amgen, Google

IT: Alex Skrenchuk, R-IT team (led by Somalee Datta)