Artificial Intelligence, Machine Learning,

and the Practice of Anesthesiology

Michael Mathis, M.D.

Assistant Professor of Anesthesiology

Affiliated Faculty, Department of Computational Medicine & Bioinformatics

Research Director, Multicenter Perioperative Outcomes Group

MICHIGAN MEDICINE
UNIVERSITY OF MICHIGAN

Disclosures

I have no personal financial, consulting, or contractual relationships with any vendor.

I am PI or Co-I on research grants:

National Institute of Health – NHLBI

Department of Defense

Overview

- Some background & context
- Near-future *examples of ML/AI* that anesthesiologists might conceivably use
- ML/AI *logistical, social, and ethical* dilemmas & what to do about them

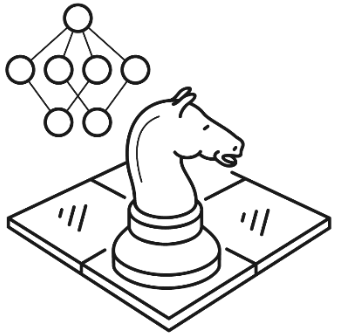
Overview

- Some background & context
- Near-future *examples of ML/AI* that anesthesiologists might conceivably use
- ML/AI *logistical, social, and ethical* dilemmas & what to do about them

…so, what are we talking about?

Types of AI

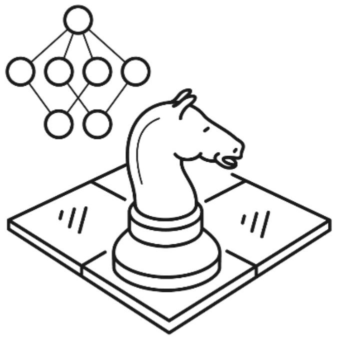
Reactive AI			
• Has no memory • Task-specific			



…so, what are we talking about?

Types of AI

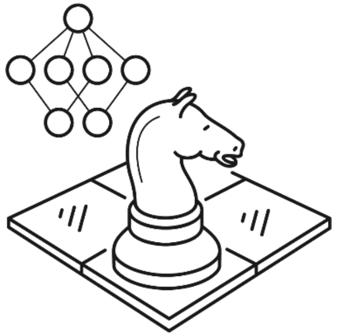
Reactive AI	Limited Memory		
• Has no memory • Task-specific	• Past experiences inform future decisions • Vulnerable to outliers or new situations		

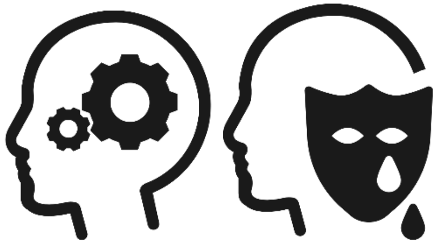


…so, what are we talking about?

Types of AI

Reactive AI	Limited Memory	Theory of Mind	
• Has no memory • Task-specific	• Past experiences inform future decisions • Vulnerable to outliers or new situations	• Understands human reasoning • Learns with fewer examples, since understands motives / intent	

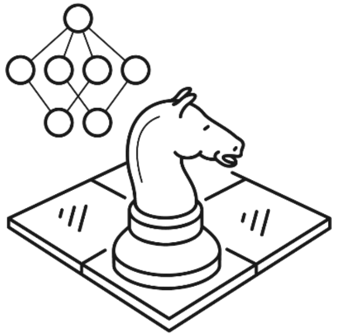


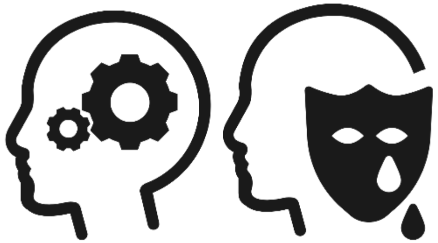


…so, what are we talking about?

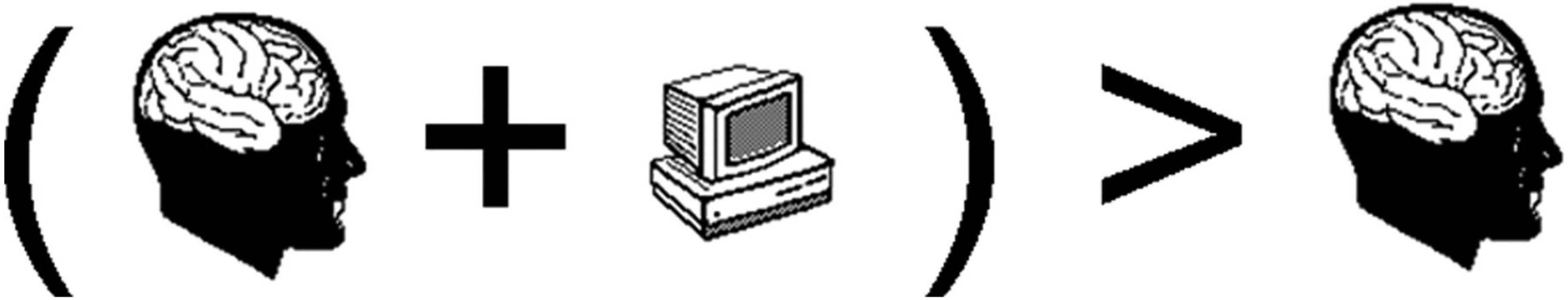
Types of AI

Reactive AI	Limited Memory	Theory of Mind	Self-aware
• Has no memory • Task-specific	• Past experiences inform future decisions • Vulnerable to outliers or new situations	• Understands human reasoning • Learns with fewer examples, since understands motives / intent	• Human-level intelligence that can bypass our own intelligence

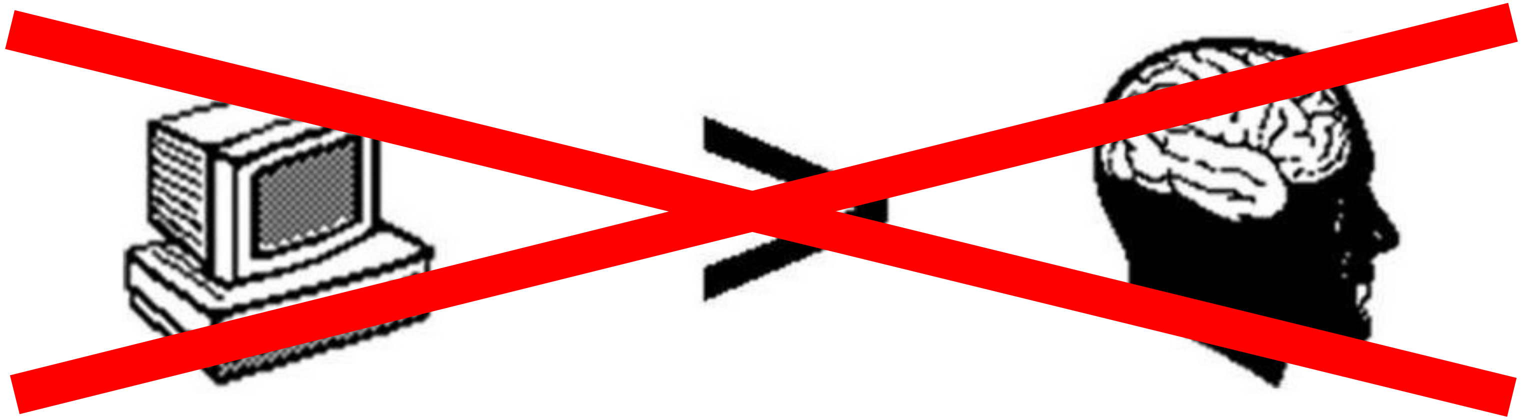




…and what is the goal?



Not:



"Fundamental Theorem" of Biomedical Informatics [1]

1. Friedman, CP. A "Fundamental Theorem" of Biomedical Informatics. *Journal of the American Medical Informatics Association.* 2009;16(2):169-170.

…and what is the goal?

	Machine Learners	Humans
Strength	Rapid, “unbiased”, accurate predictions	Understand implications of clinical decisions
Weakness	Lack transparency Lack clinical judgment	High cost, slow

…and why care?

…and finally, why is AI/ML so hard in healthcare?

Mistakes matter

- Consequences are not just annoying, but dangerous
- Mistakes by a human preferred to mistakes by computer

Being "right" isn't enough

…Need to be *transparent* and *justified*

Nobody likes to share

(and often can't, even if we wanted to)

Overview

- Some background & context
- Near-future *examples of ML/AI* that anesthesiologists might conceivably use
- ML/AI *logistical, social, and ethical* dilemmas & what to do about them

Overview

- Some background & context
- Near-future *examples of ML/AI* that anesthesiologists might conceivably use
- ML/AI *logistical, social, and ethical* dilemmas & what to do about them

ML/AI example: ICU Deterioration Prediction

ANNALS OF THE AMERICAN THORACIC SOCIETY®

Validating a Widely Implemented Deterioration Index Model Among Hospitalized COVID-19 Patients

Singh K, Valley TS, Tang S, Li BY, Kamran F, Sjoding MW, Wiens J, Otles E, Donnelly JP, Wei MY, McBride JP, Cao J, Penoza C, Ayanian JZ, Nallamothu BK: Validating a Widely Implemented Deterioration Index Model Among Hospitalized COVID-19 Patients. AnnalsATS *In Press Dec 2020.*

ML/AI example: ICU Deterioration Prediction

- **Goal:** Early detection of patient deterioration (ICU transfer, mechanical ventilation, or death)
- **Cohort:** UMich COVID-19 patients admitted to non-ICU
- **Data Collected:**
  - **Demographics** – Age, gender, race
  - **Vitals** – SBP, HR, RR, SpO_2
  - **RN assessments** – GCS, ECG rhythm, O_2 requirement
  - **Lab values** – CBC, chemistry, ABG

ML/AI example: ICU Deterioration Prediction

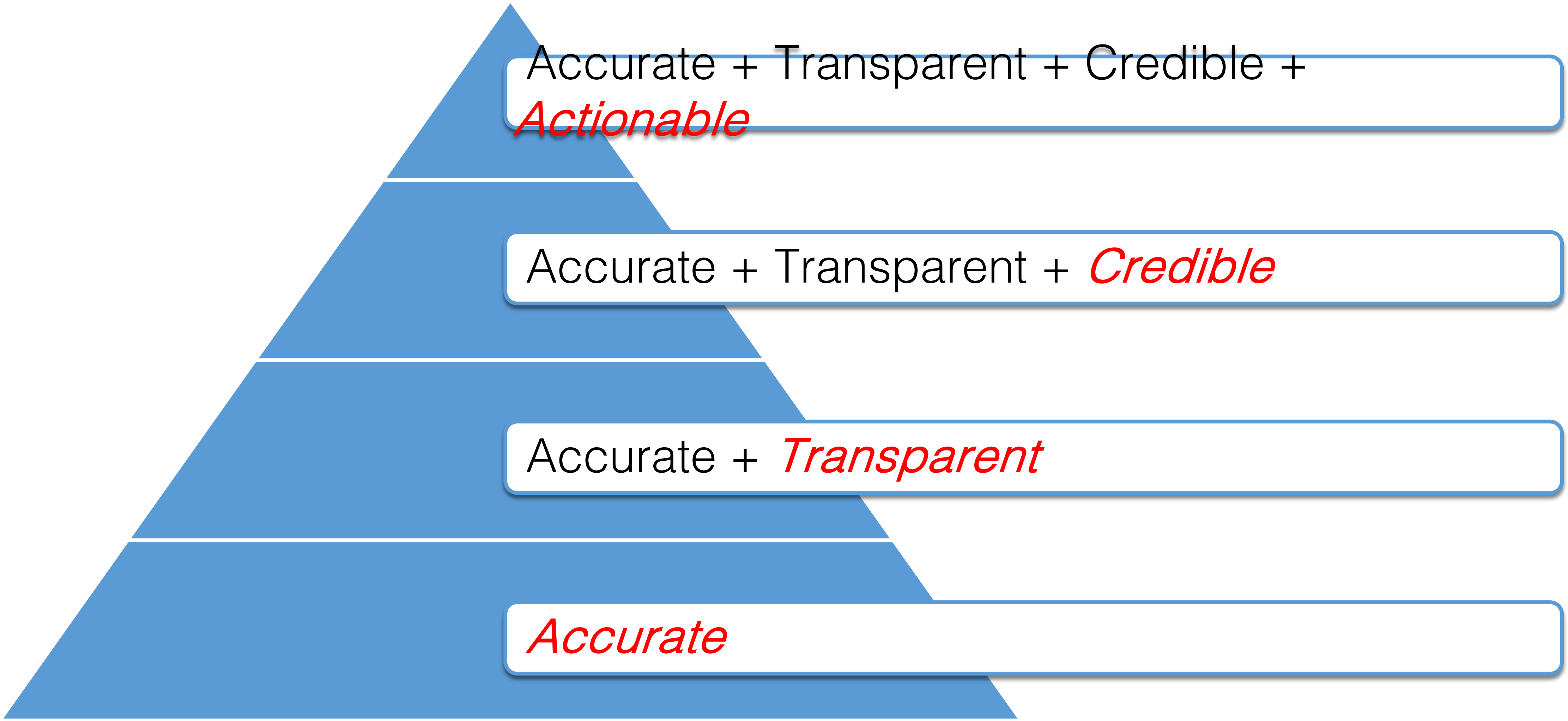
- **Model Output:** Score 0-100, representing likelihood of deterioration

- Algorithm performance:
    - AUC 0.76 (95% CI 0.68-0.84)

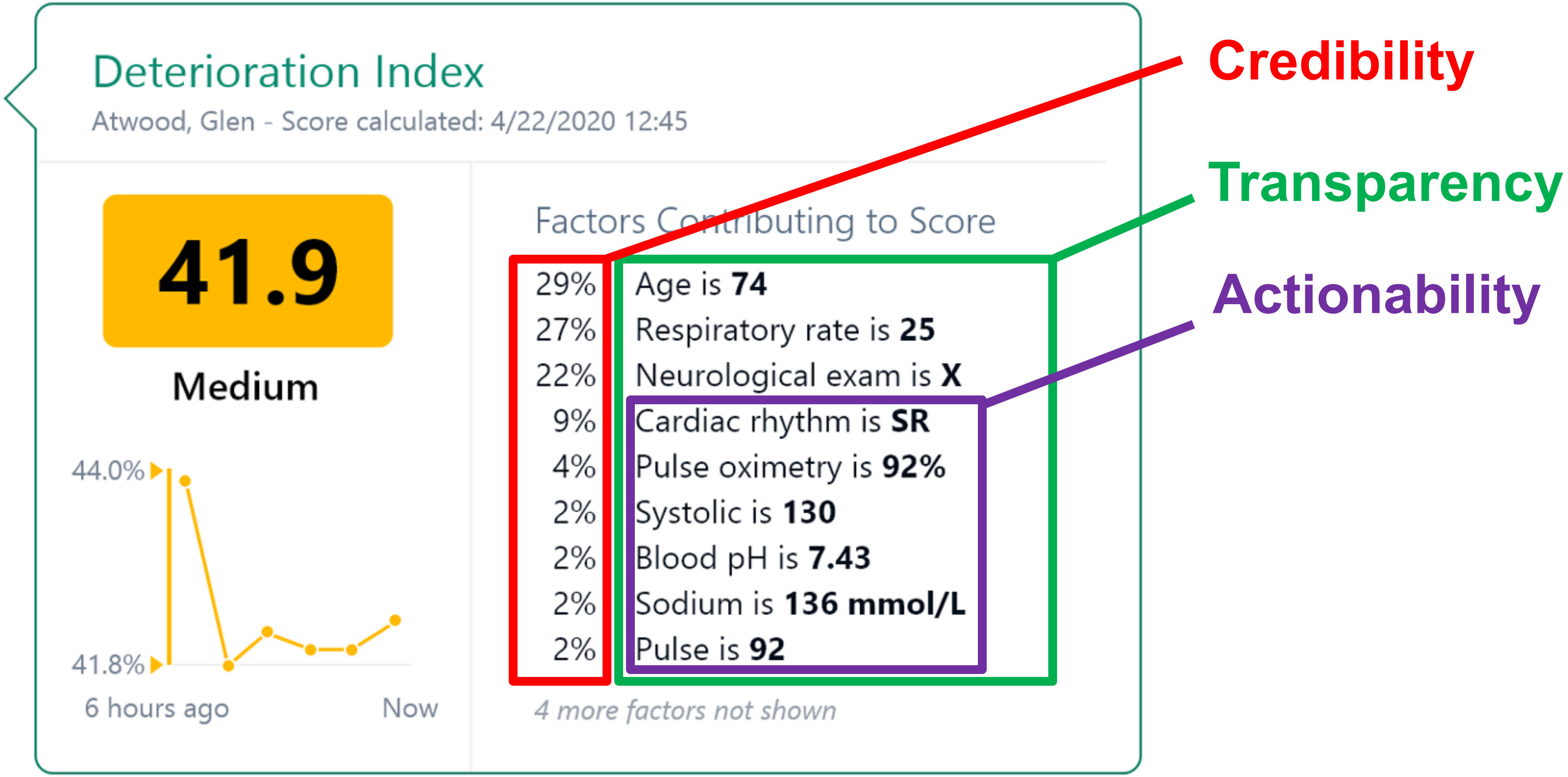
… so the model works *modestly* well, but:

- How does it work?
- Why should we trust (or not trust) it? (And in what situations?)
- What should we do with the information?

Dilemma: Model Trust & Utility

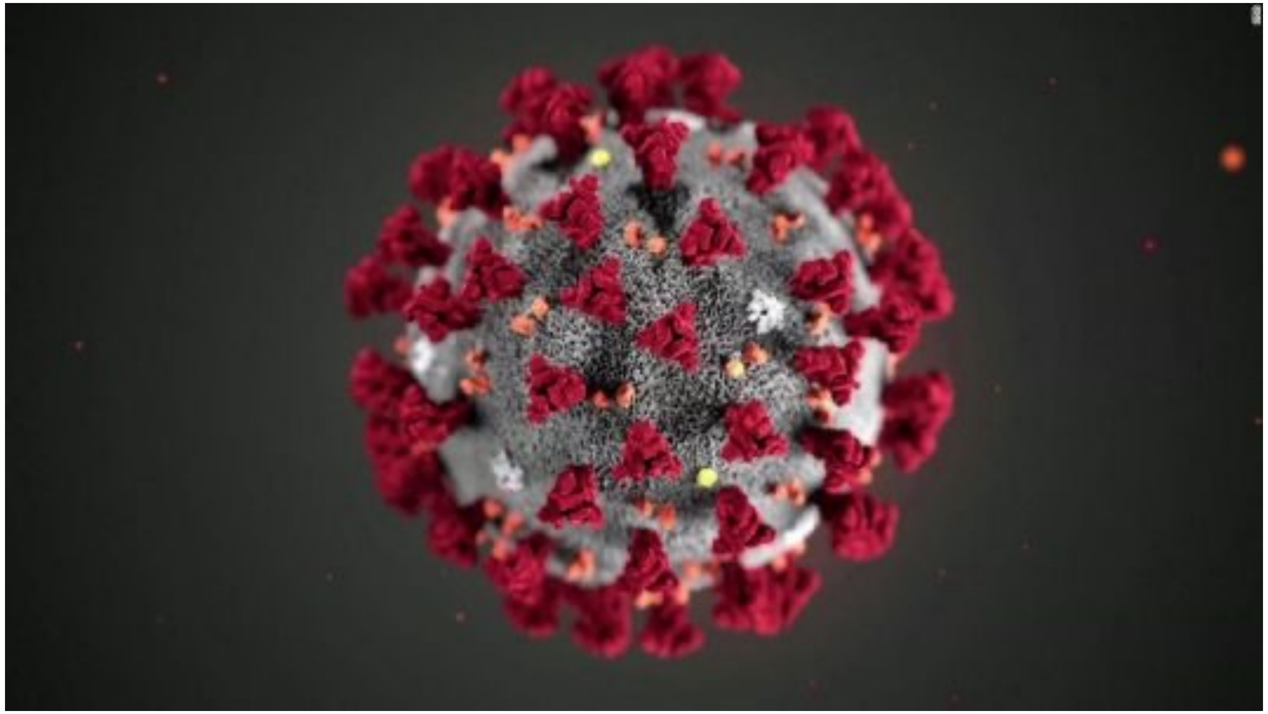


Dilemma: Model Trust & Utility



Dilemma: *Dataset Shift*

Underperformance of a ML system due to ***evolving mismatch*** *between training data and test data*



Dilemma: *Dataset Shift*

Requires both vigilant frontline clinicians, and oversight from AI governance teams[1]

1. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, Kohane IS, Saria S: The Clinician and Dataset Shift in Artificial Intelligence. N Engl J Med 2021; 385:283–61

ML/AI example: Post-induction hypotension prediction

ANESTHESIOLOGY

Supervised Machine-learning Predictive Analytics for Prediction of Postinduction Hypotension

Kendale S, Kulkarni P, Rosenberg AD, Wang J: Supervised Machine-learning Predictive Analytics for Prediction of Postinduction Hypotension. Anesthesiology 2018; 129:675–88

ML/AI example: Post-induction hypotension prediction

- **Goal:** Use data available *prior to induction* to predict likelihood of post-induction hypotension
- **Dataset:** NYU surgical patients >12 years old

Medical History **Data**

- Demographics
- Comorbidities
- Home Meds

Perioperative Data

- ASA Status
- Intraoperative Meds
- Intraoperative Physiologic

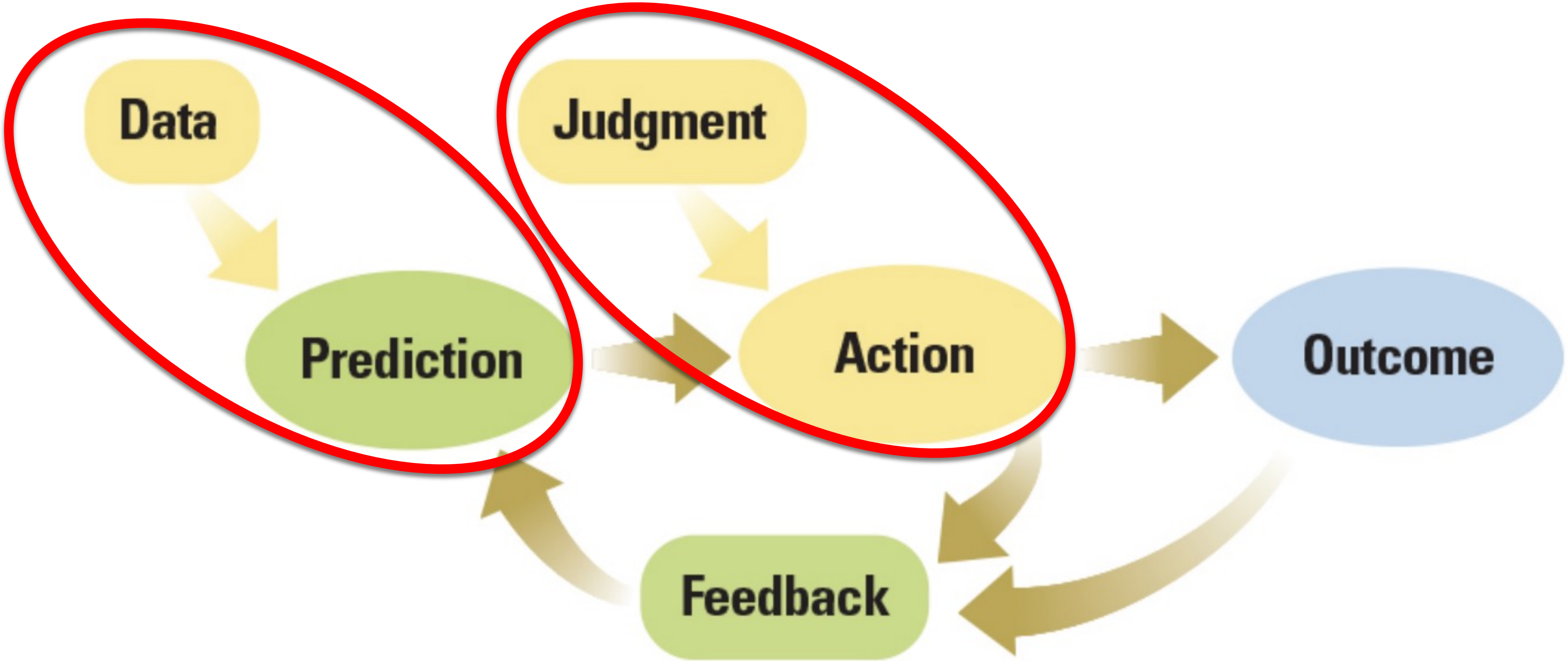
ML/AI example: Post-induction hypotension prediction

- In a test dataset, the best prediction model yielded modest performance 10 minutes prior to induction:

  AUC: 0.74 (95% CI 0.72-0.77)

- …so what do we do with this information?

Dilemma: Quantifying Clinical Judgment



Dilemma: Quantifying Clinical Judgment

- 74yo ASA 4 male with carotid stenosis, severe COPD and Stage 3 CKD undergoing colectomy

ML Prediction algorithm

Post-induction Hypotension probability: *98%*

ANESTHESIOLOGY

Supervised Machine-learning Predictive Analytics for Prediction of Postinduction Hypotension

Dilemma: Quantifying Clinical Judgment

- 74yo ASA 4 male with carotid stenosis, severe COPD and Stage 3 CKD undergoing colectomy
- Post-induction Hypotension probability: *98%*

Dilemma: Quantifying Clinical Judgment

- 74yo ASA 4 male with carotid stenosis, severe COPD and Stage 3 CKD undergoing colectomy
- Post-induction Hypotension probability: *98%*

Action	Outcome: Stroke	Outcome: Pulm Complication	Outcome: AKI
Do Nothing	1%	10%	10%

Dilemma: Quantifying Clinical Judgment

- 74yo ASA 4 male with carotid stenosis, severe COPD and Stage 3 CKD undergoing colectomy
- Post-induction Hypotension probability: *98%*

Action	Outcome: Stroke	Outcome: Pulm Complication	Outcome: AKI
Do Nothing	1%	10%	10%
Fluid Bolus	0.5%	20%	5%

Dilemma: Quantifying Clinical Judgment

- 74yo ASA 4 male with carotid stenosis, severe COPD and Stage 3 CKD undergoing colectomy
- Post-induction Hypotension probability: *98%*

Action	**Outcome: Stroke**	**Outcome: Pulm Complication**	**Outcome: AKI**
Do Nothing	1%	10%	10%
Fluid Bolus	0.5%	20%	5%
Vasopressor	0.5%	10%	20%

Dilemma: Quantifying Clinical Judgment

Judgment to be made…
…what matters the most?
…and by ***how much more***?

Action	**Outcome: Stroke**	**Outcome: Pulm Complication**	**Outcome: AKI**
Do Nothing	1%	10%	10%
Fluid Bolus	0.5%	20%	5%
Vasopressor	0.5%	10%	20%

Dilemma: Quantifying Clinical Judgment

	Judgments for Clinician Action = "Policy"		
Action	**Outcome: Stroke**	**Outcome: Pulm Complication**	**Outcome: AKI**
Do Nothing	1%	10%	10%
Fluid Bolus	0.5%	20%	5%
Vasopressor	0.5%	10%	20%

Dilemma: Quantifying Clinical Judgment

	Burden of Stroke:	Burden of Pulm Comp:	Burden of AKI:
	100	***2***	***1***

Judgments for Clinician Action = "Policy"

Action	**Outcome: Stroke**	**Outcome: Pulm Complication**	**Outcome: AKI**
Do Nothing	1%	10%	10%
Fluid Bolus	0.5%	20%	5%
Vasopressor	0.5%	10%	20%

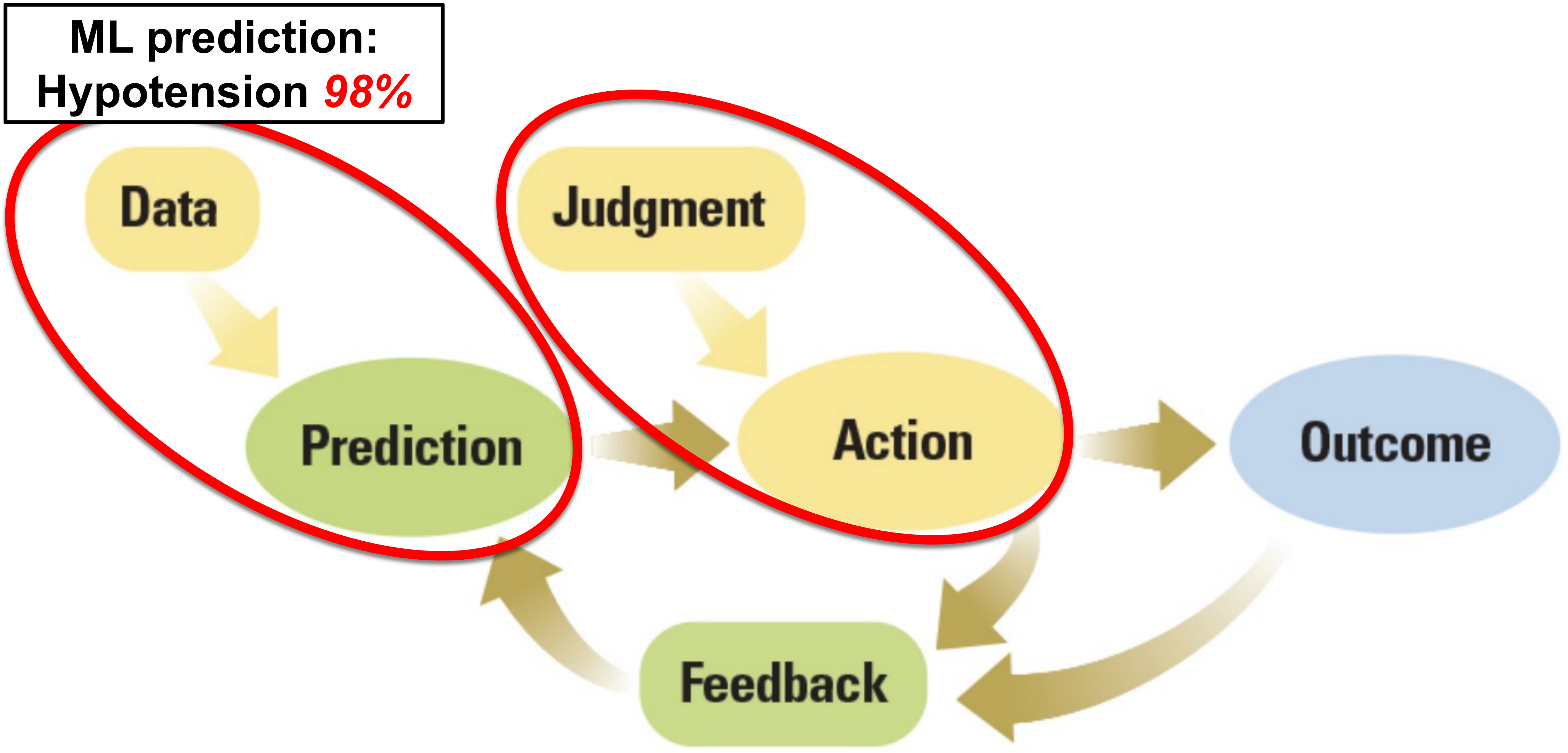
Dilemma: Quantifying Clinical Judgment

Burden of Stroke:	Burden of Pulm Comp:	Burden of AKI:
100	*2*	*1*

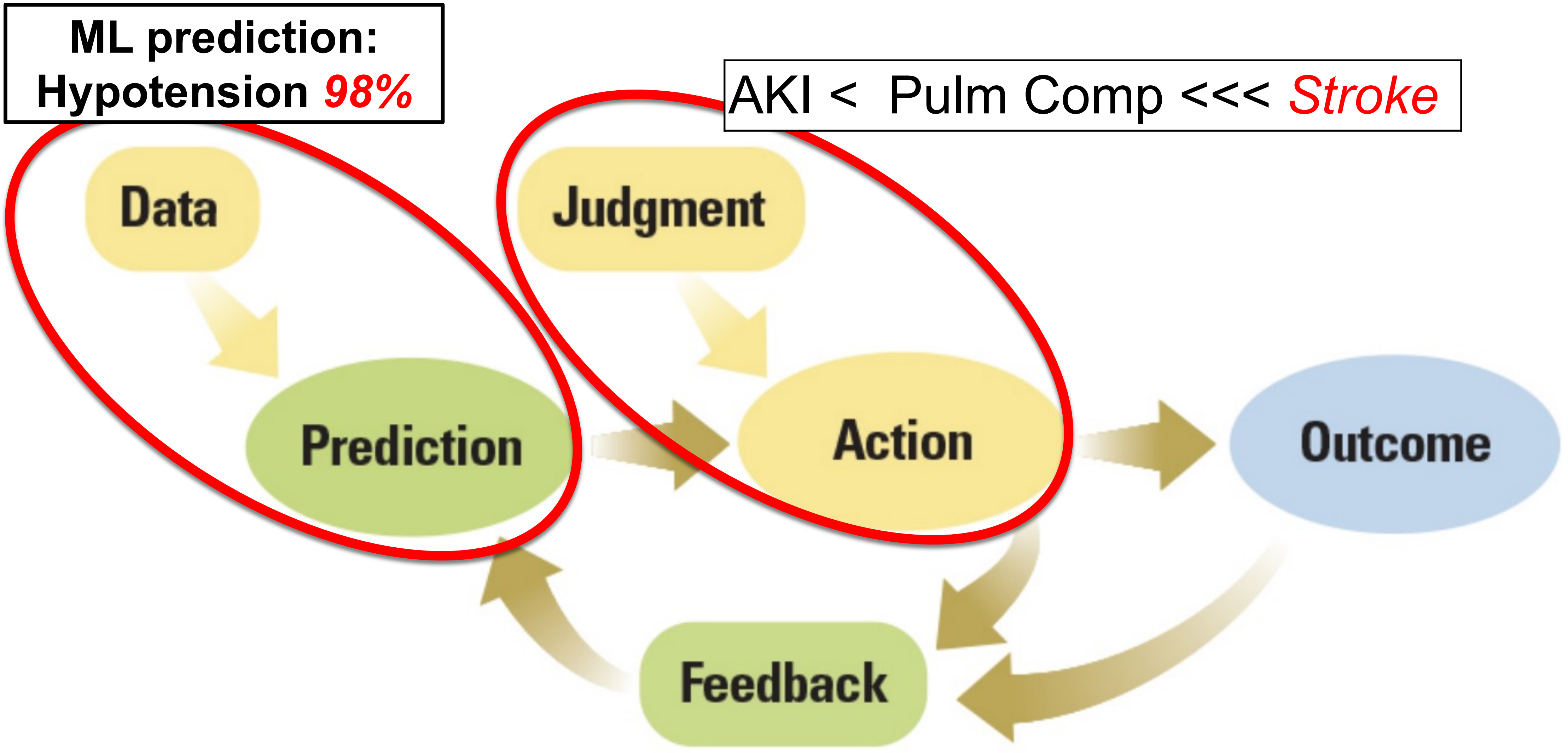
Who decides this?

Ac				d
Do Nothing	1% * *100*	10% * *2*	10% * *1*	*1.30*
Fluid Bolus	0.5% * *100*	20% * *2*	5% * *1*	*0.95*
Vasopressor	0.5% * *100*	10% * *2*	20% * *1*	*0.90*

Dilemma: Quantifying Clinical Judgment



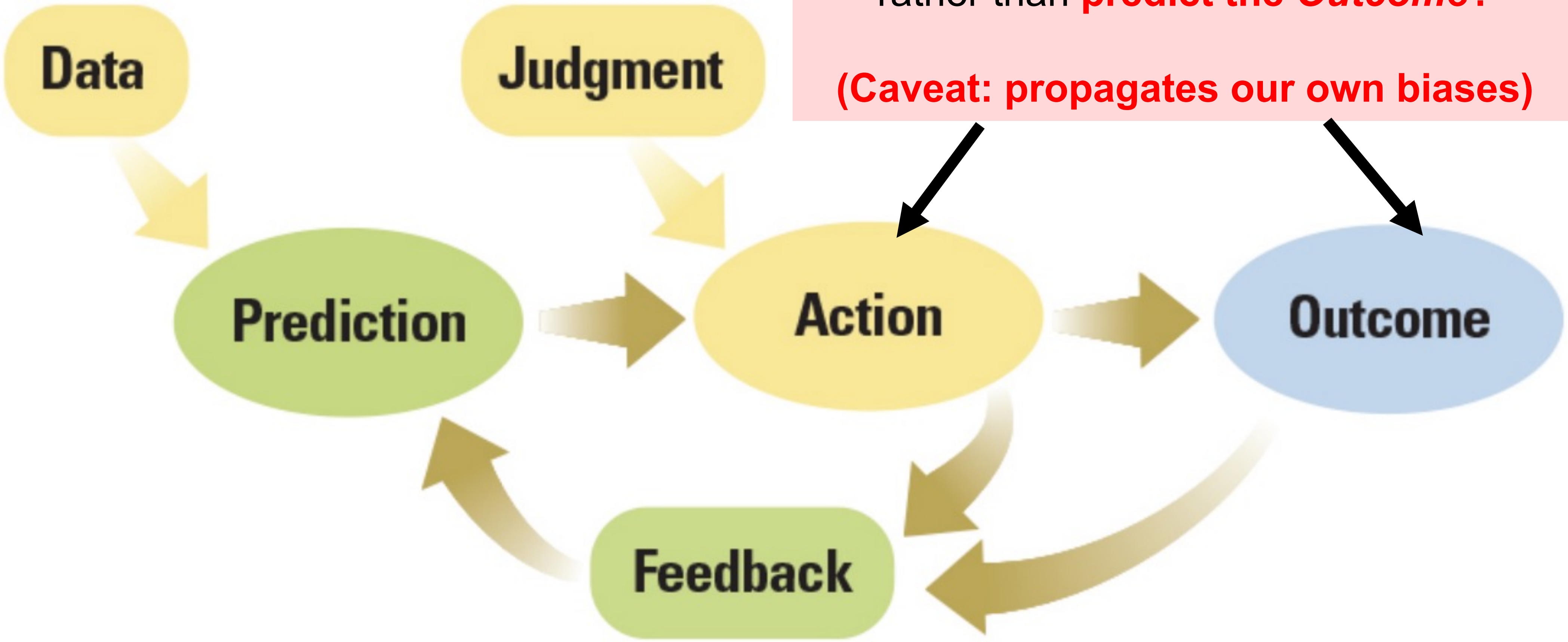
Dilemma: Quantifying Clinical Judgment



Dilemma: Quantifying Clinical Judgment

Predict the *Action* (to infer judgment)
rather than **predict the *Outcome*?**

(Caveat: propagates our own biases)



Dilemma: Quantifying Clinical Judgment

Take-Home Points

- Machine learning for perioperative care is *conceptually feasible*

Take-Home Points

- Machine learning for perioperative care is *conceptually feasible*

- Next steps are to ensure ML algorithms are *generalizable*, *safe*, *unbiased*, and *clinically actionable*

What you can do today

- Develop literacy in AI/ML methods/interpretation
- Know the right problems for AI/ML to solve
- Understand the AI/ML model inputs
- Seek transparency and actionability
- Learn your biases …(before a ML model does)

Thank you