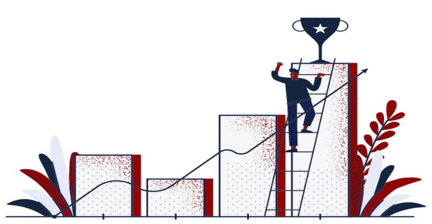
The F-ai-rest of them all.

How To Do Healthy Machine Learning in Health.

Dr. Marzyeh Ghassemi
MIT IMES/EECS.CSAIL
CIFAR AI Chair, Azrieli Global Scholar



what **models** are **healthy**?

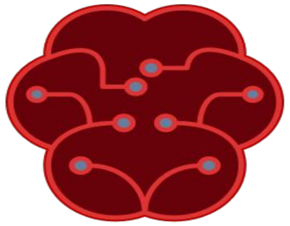
what **healthcare** is **healthy**?

what **behaviors** are **healthy**?




Machine Learning in Health 101

- We collect **data** on those who are **sick.**



- We predict when **bad events** or **outcomes**, e.g., worsening sickness, happens.

SOTA Methods At/Above Human Performance

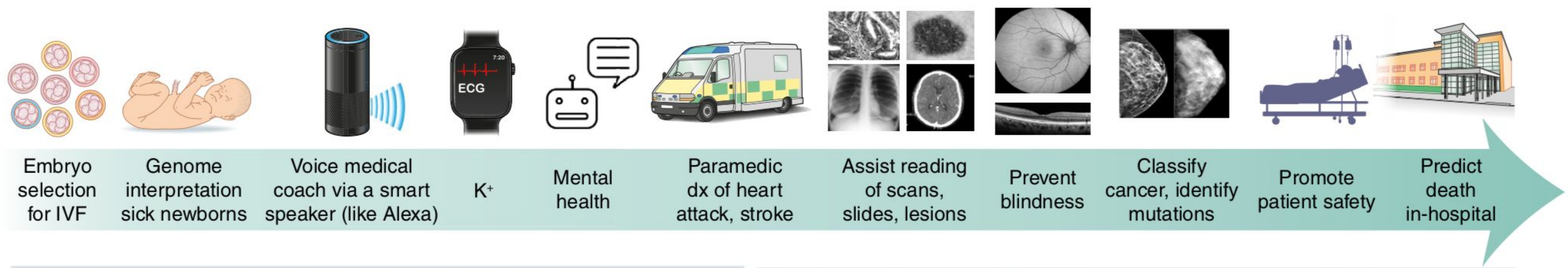


Table 3 | Selected reports of machine- and deep-learning algorithms to predict clinical outcomes and related parameters

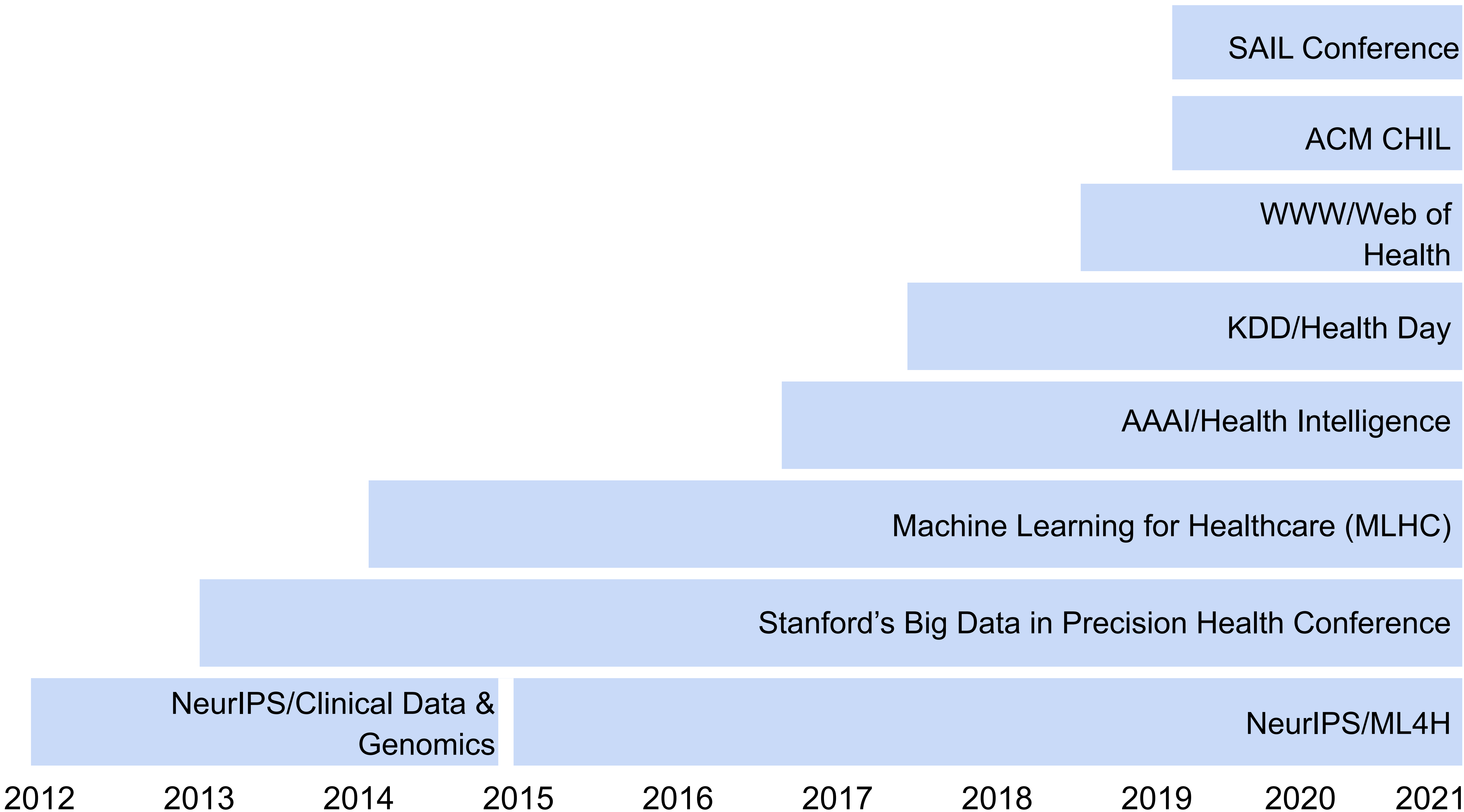
Prediction	*n*	AUC	Publication (Reference number)
In-hospital mortality, unplanned readmission, prolonged LOS, final discharge diagnosis	216,221	$0.93^{*}0.75^{+}0.85^{\#}$	Rajkomar et al.[96]
All-cause 3-12 month mortality	221,284	$0.93^{\wedge}$	Avati et al.[91]
Readmission	1,068	0.78	Shameer et al.[106]
Sepsis	230,936	0.67	Horng et al.[102]
Septic shock	16,234	0.83	Henry et al.[103]
Severe sepsis	203,000	$0.85^{@}$	Culliton et al.[104]
Clostridium difficile infection	256,732	0.82^{++}	Oh et al.[93]
Developing diseases	704,587	range	Miotto et al.[97]
Diagnosis	18,590	0.96	Yang et al.[90]
Dementia	76,367	0.91	Cleret de Langavant et al.[92]
Alzheimer's Disease (+ amyloid imaging)	273	0.91	Mathotaarachchi et al.[98]
Mortality after cancer chemotherapy	26,946	0.94	Elfiky et al.[95]
Disease onset for 133 conditions	298,000	range	Razavian et al.[105]
Suicide	5,543	0.84	Walsh et al.[86]
Delirium	18,223	0.68	Wong et al.[100]

LOS, length of stay; *n*, number of patients (training+validation datasets). For AUC values: *, in-hospital mortality; +, unplanned readmission; #, prolonged LOS; ^, all patients; @, structured+unstructured data; ++, for University of Michigan site.

Source: **High-performance medicine: the convergence of human and artificial intelligence** Eric Topol, Nature Medicine Jan 2019

Figure: Debbie Maizels / Springer Nature

ML Researchers Have Really Engaged



Models Are Regulated Advice-Givers

FDA Cleared AI Algorithms

Our list of FDA cleared AI algorithms provides valuable details on each model, bringing all of the relevant information together for easy access. Convenient summaries for each algorithm include model manufacturer, FDA product code, body ar predicate devices, product testing and evaluation related to product performance, and clinical validation. Our Define-AI use cases match many of the models and those are listed under Related Use Cases. For other details, clicking on the mode directly to the FDA summary.

Check back regularly to see which new algorithms are available and have been added to the list. Send information on AI algorithms that are not listed and report missing information to DSI@acr.org.

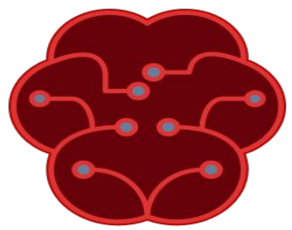
Search | Company | Subspeciality | Body Area | Modality | Date Cleared

Product	Company	Subspeciality	Body Area	Modality	Date Cleared
Transpara 1.7.0	ScreenPoint Medical B.V.	Women's Imaging	Breast	MAM	06/02/2021
CINA CHEST	AVICENNA.AI	Chest Imaging	Pulmonary Arteries	CT	05/19/2021
Overjet Dental Assist	Overjet, Inc.	Dental Imaging	Teeth	XRAY	05/19/2021
MEDO- Thyroid	Medo.AI	Chest Imaging	Thyroid	US	04/23/2021
Saige-Q	DeepHealth	Women's Imaging	Breast	MAM	04/16/2021
syngo.CT Lung CAD (VD20)	Siemens Healthineers	Chest Imaging	Lung	CT	03/31/2021
Viz ICH	Viz. ai, inc.	Neuroradiology	Brain	CT	03/23/2021
Vbrain	Vysioneer Inc.	Neuroradiology	Brain	MR	03/19/2021
Imbio RV/LV Software	Imbio LLC	Cardiac Imaging	Heart	CT	03/09/2021
Optellum Virtual Nodule Clinic, Optellum Software, Optellum Platform	Optellum Ltd.	Chest Imaging	Lung	CT	03/05/2021
NinesMeasure	Nines, Inc.	Chest Imaging	Chest	CT	02/25/2021
Veolity	MeVis Medical Solutions AG	Chest Imaging	Chest	CT	02/23/2021
Lvivo Software Application	DiA Imaging Analysis Ltd	Cardiac Imaging,Abdominal Imaging	Heart,Head	US	02/05/2021
qp-Prostate	Quibim	Abdominal Imaging	Prostate	MR	02/04/2021
Visage Breast Density	Visage Imaging GmbH	Women's Imaging	Breast	MAM	01/29/2021
uAI EasyTriage-Rib	Shanghai United Imaging Intelligence Co., Ltd.	Chest Imaging	Chest	CT	01/15/2021
HearFlow Analysis	HeartFlow, Inc.	Cardiac Imaging	Coronary Arteries	CT	01/08/2021
BrainInsight	Hyperfine Research, Inc.	Neuroradiology	Brain	MR	01/07/2021

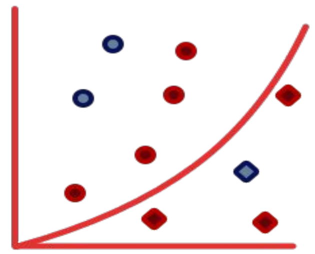
Source: **https://models.acrdsi.org/** August 31, 2021

Machine Learning in Health 101

- We collect **data** on those who are **sick.**

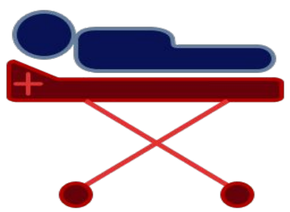


- We predict when **bad events** or **outcomes**, e.g., worsening sickness, happens.

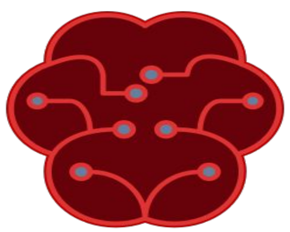


- Bad events in sick people is actually **anomaly detection** in **anomalous data**.

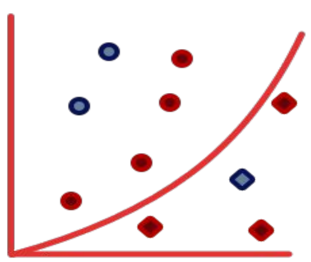
Machine Learning in Health 101



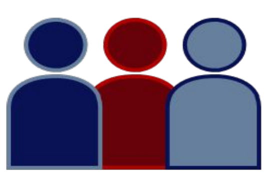
- We collect **data** on those who are **sick.**



- We predict when **bad events** or **outcomes**, e.g., worsening sickness, happens.



- Bad events in sick people is actually **anomaly detection** in **anomalous data**.

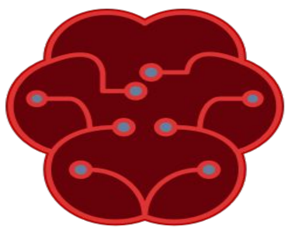


- We have no idea what it means for a **diverse population** to be **healthy**.

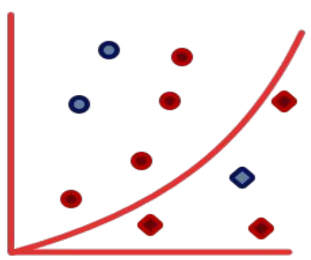
6

Machine Learning in Health 101

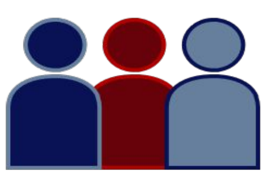
- We collect **data** on those who are **sick.**



- We predict when **bad events** or **outcomes**, e.g., worsening sickness, happens.



- Bad events in sick people is actually **anomaly detection** in **anomalous data**.



- We have no idea what it means for a **diverse population** to be **healthy**.

What exactly are we learning?

Healthy Machine Learning in Health

what **models** are **healthy**?

what **healthcare** is **healthy**?

what **behaviors** are **healthy**?

Creating actionable insights in human health.

8

Healthy Machine Learning in Health

what **models** are **healthy**?

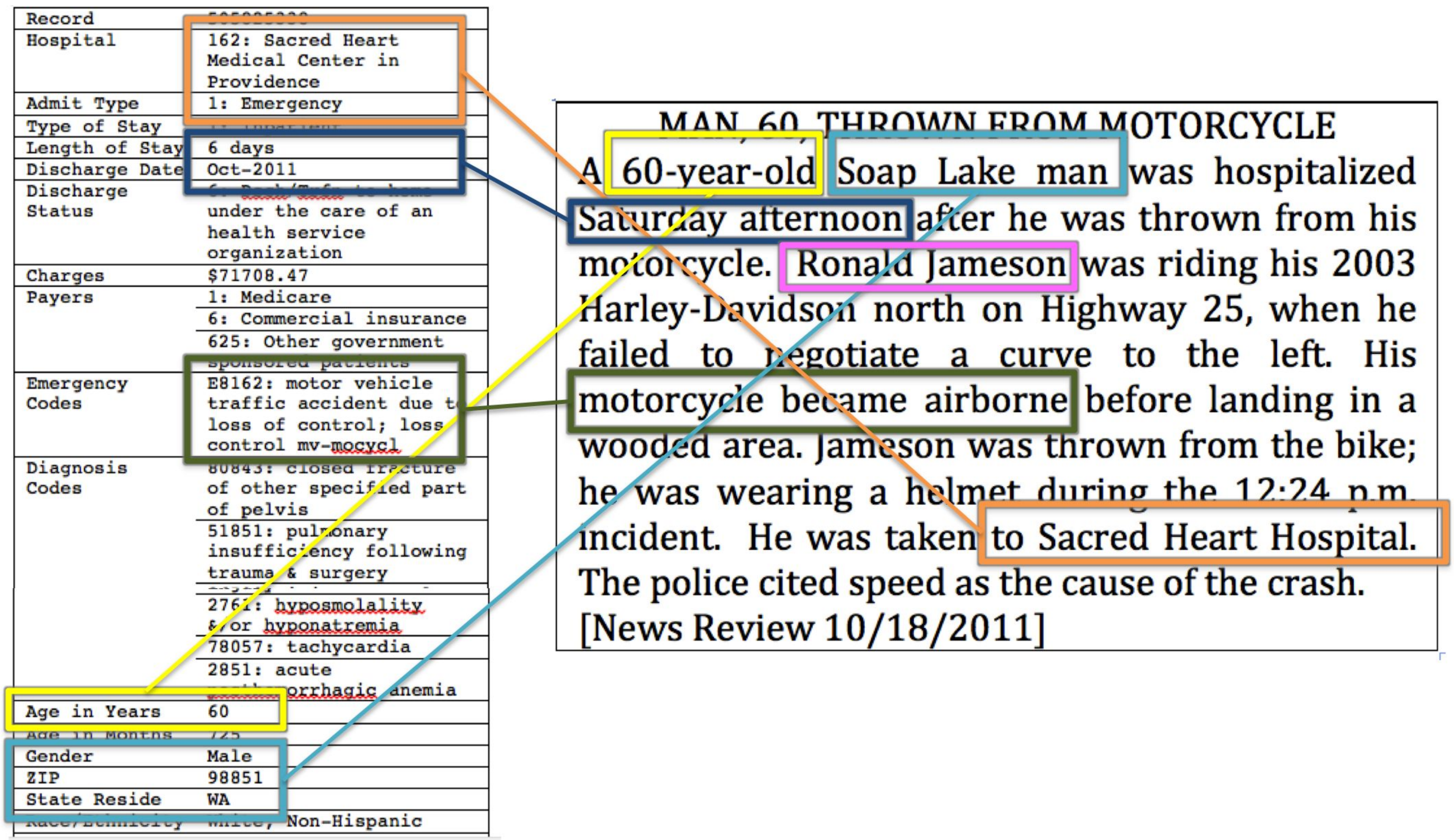
what **healthcare** is **healthy**?

what **behaviors** are **healthy**?

Creating actionable insights in human health.

Privacy in Clinical Prediction Models

- Anonymization is not robust to linkage.



Record	[illegible]
Hospital	162: Sacred Heart Medical Center in Providence
Admit Type	1: Emergency
Type of Stay	[illegible]
Length of Stay	6 days
Discharge Date	Oct-2011
Discharge Status	[illegible] under the care of an health service organization
Charges	$71708.47
Payers	1: Medicare
	6: Commercial insurance
	625: Other government sponsored patients
Emergency Codes	E8162: motor vehicle traffic accident due to loss of control; loss control mv-mocycl
Diagnosis Codes	80843: closed fracture of other specified part of pelvis
	51851: pulmonary insufficiency following trauma & surgery
	2761: hyposmolality &/or hyponatremia
	78057: tachycardia
	2851: acute posthemorrhagic anemia
Age in Years	60
Age in Months	725
Gender	Male
ZIP	98851
State Reside	WA
Race/Ethnicity	White, Non-Hispanic

MAN, 60, THROWN FROM MOTORCYCLE

A 60-year-old Soap Lake man was hospitalized Saturday afternoon after he was thrown from his motorcycle. Ronald Jameson was riding his 2003 Harley-Davidson north on Highway 25, when he failed to negotiate a curve to the left. His motorcycle became airborne before landing in a wooded area. Jameson was thrown from the bike; he was wearing a helmet during the 12:24 p.m. incident. He was taken to Sacred Heart Hospital. The police cited speed as the cause of the crash. [News Review 10/18/2011]

Why Differential Privacy?

- In healthcare settings, it is crucial that we provide the same level of privacy protection for all individuals.

Sumana is in dataset A: gender, race, age, and zip code.

These properties alone make her identifiable to an adversary who can access the data, or the outputs of a model trained on the data.

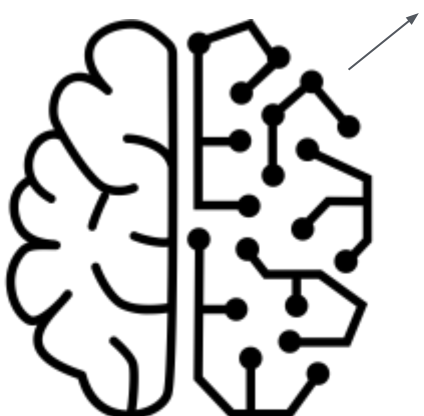
Differential privacy protects those with combination of attributes that are uniquely identifiable.

11

Differential Privacy in Yearly Mortality Prediction

- Evaluate year-to-year performance with privacy guarantee $\epsilon = \frac{\sqrt{2log(1/\delta)}}{\sigma^2 C^2}$

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq \exp(\varepsilon) \Pr[\mathcal{M}(y) \in \mathcal{S}] + \delta,$$



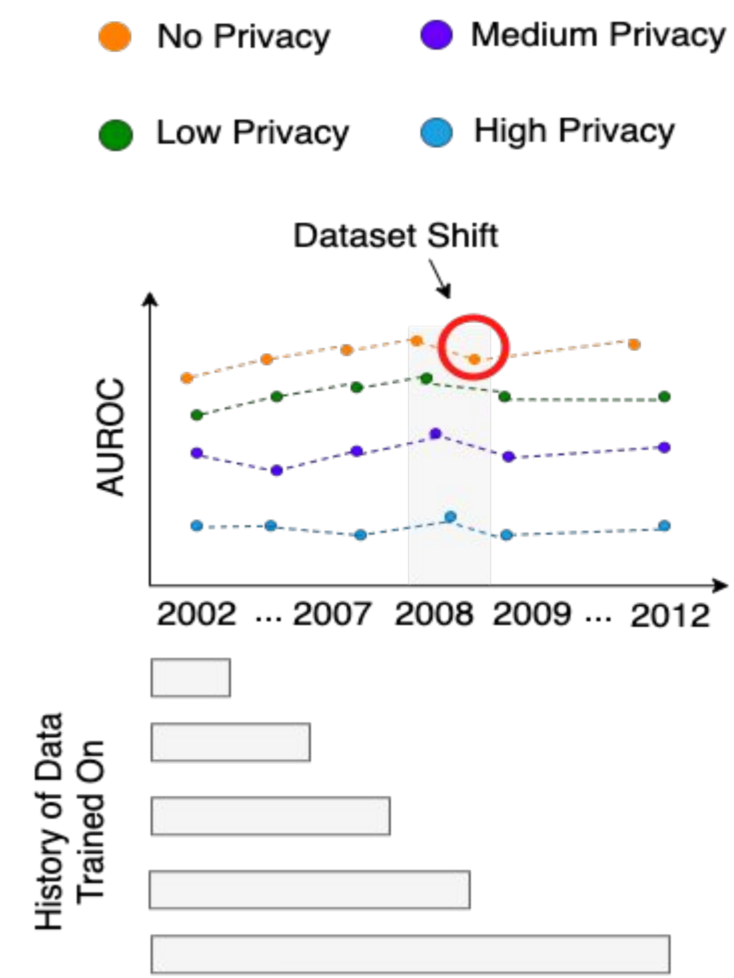
$$\mathcal{L}(\theta(x), y)$$

Task	Prediction	True Label
Mortality	0	1

$$g = \nabla_\theta \mathcal{L}(\theta(x), y)$$

Update Model

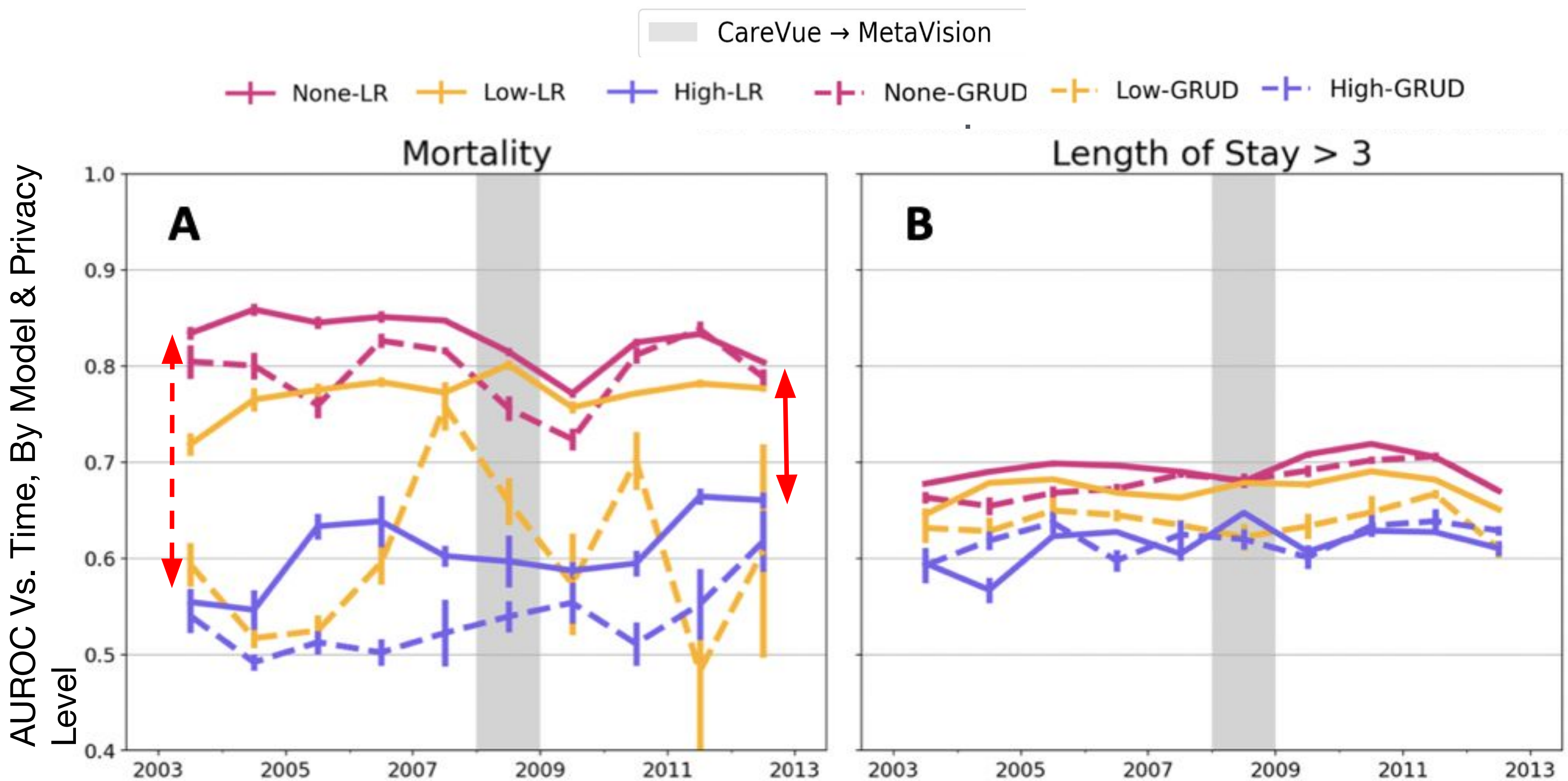
$$\frac{\begin{bmatrix} 0.06 & -1.13 \\ 1.10 & -0.56 \end{bmatrix} + \mathcal{N}(0, \sigma^2 \cdot C^2)}{max\left(1, \frac{||\nabla_\theta g||}{C}\right)}$$



[1] Suriyakumar, Papernot, Goldenberg, Ghassemi. "Chasing Your Long Tails: Differentially Private Prediction in Health Care Settings." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (*FAccT '21*).

Differential Privacy vs. Utility Trade-off

- What price are we willing to pay for differential privacy?



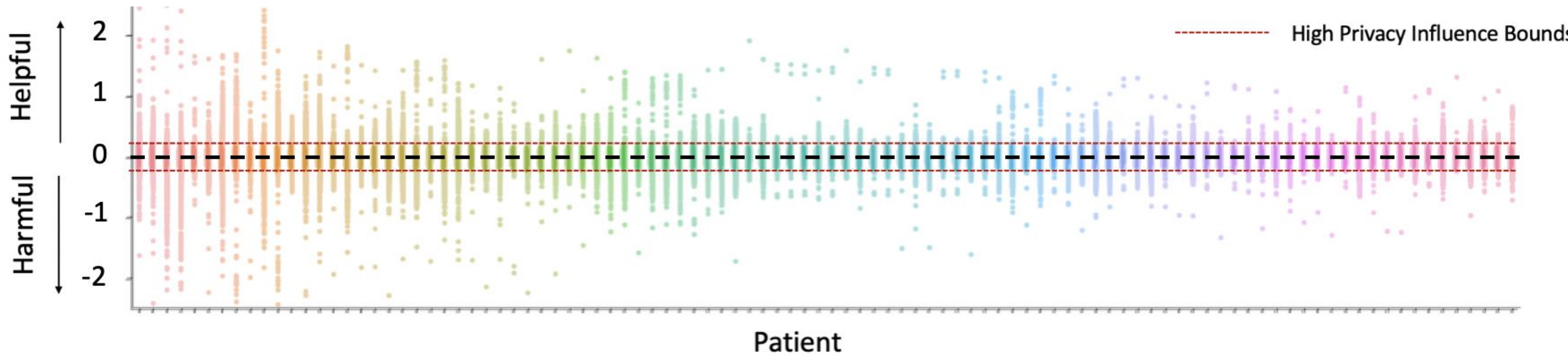
Large loss of performance for high privacy models in mortality.

Less severe drops in tasks with lower initial performance.

[1] Suriyakumar, Papernot, Goldenberg, Ghassemi. "Chasing Your Long Tails: Differentially Private Prediction in Health Care Settings." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (*FAccT '21*).

ML is Built on Finding and Enforcing **Similarity**

- Training "data" loses predictive **influence** on test with more privacy.



- Some patients lose **more influence** than others.

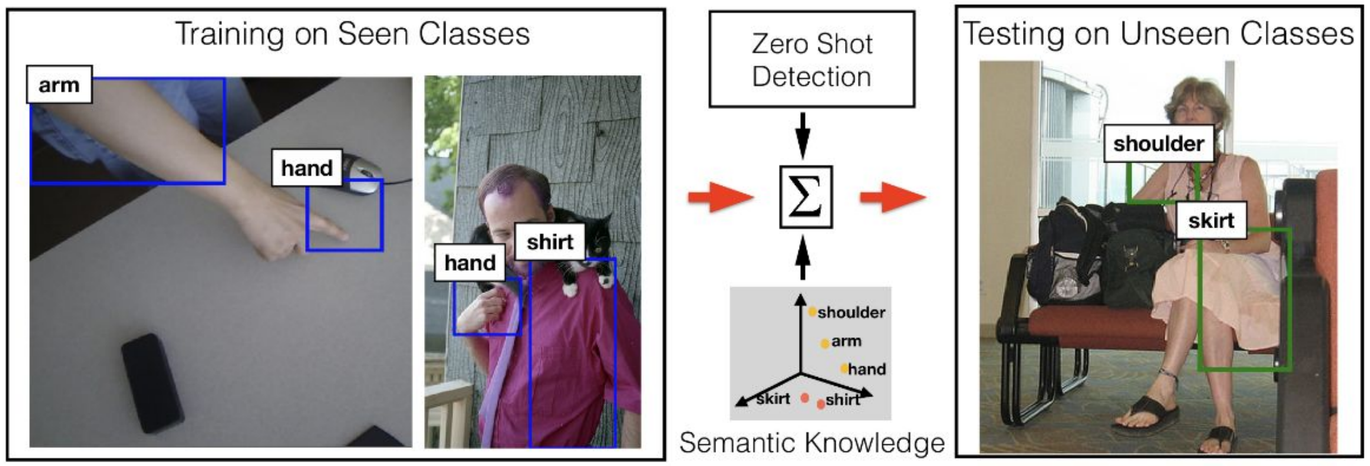
Black Test Patients				
Privacy Level	Average White Influence	Average Black Influence	Most Helpful Ethnicity	Most Harmful Ethnicity
None	0.48 ± 1.39	0.44 ± 2.19	**Black**	White
Low	-0.23 ± 0.75	-0.03 ± 0.18	White	White
High	-0.40 ± 4.10	0.12 ± 1.45	**White**	White

Adding **privacy changes** the most **helpful group training data** from Black patients to White patients for **Black test patients**.

[1] Suriyakumar, Papernot, Goldenberg, Ghassemi. "Chasing Your Long Tails: Differentially Private Prediction in Health Care Settings." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (*FAccT '21*).

Finding and Enforcing **Similarity**

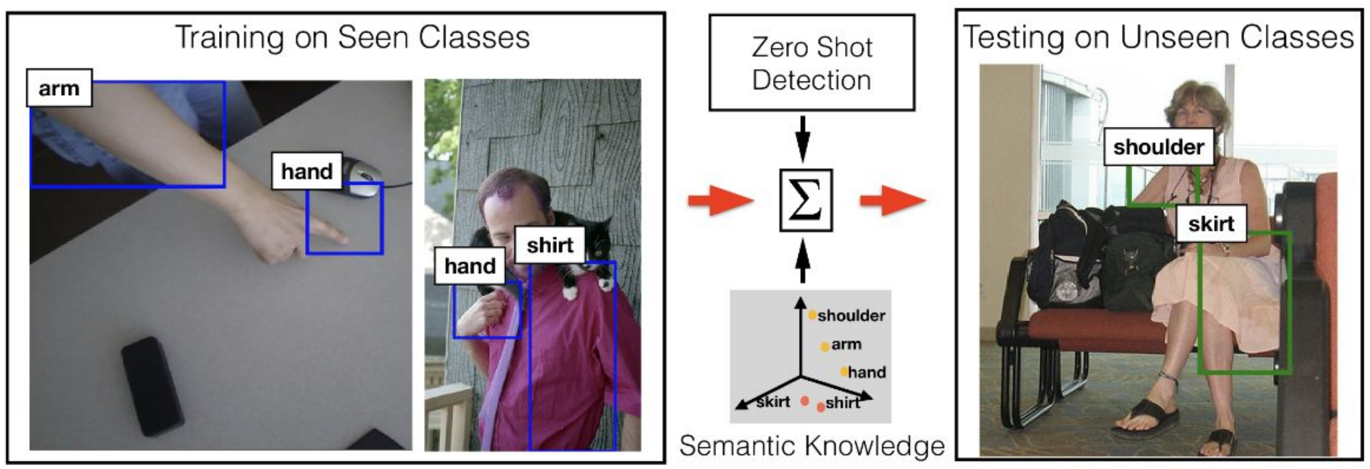
- Machine Learning is built on finding **patterns** in data, extending them, and removing **outliers**.



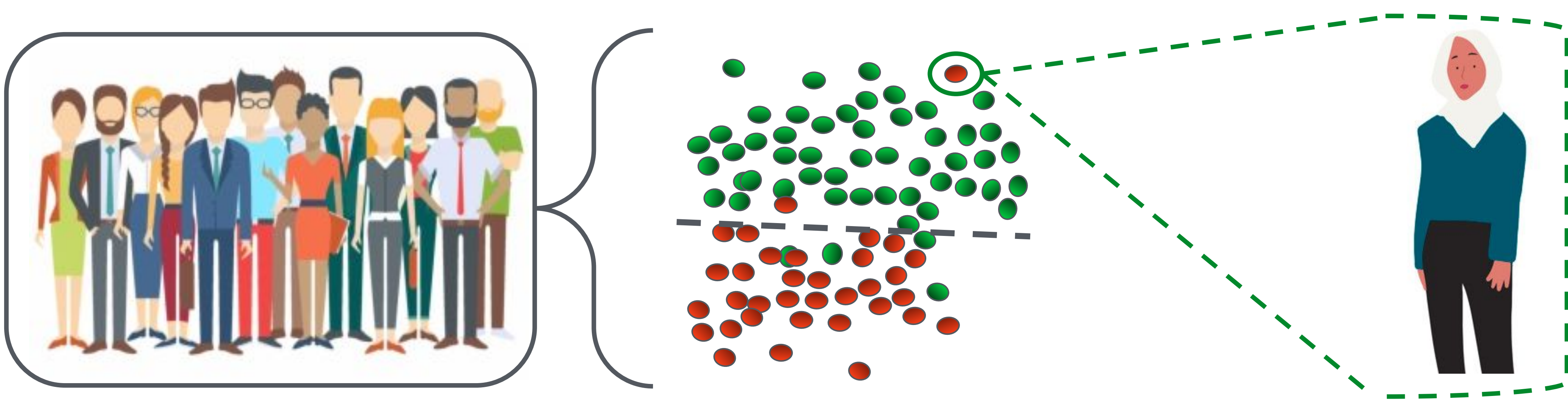
Bansal, Ankan, et al. "Zero-shot object detection." *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.

<u>Finding</u> and <u>Enforcing</u> **Similarity**

- Machine Learning is built on finding **patterns** in data, extending them, and removing **outliers**.



- What does it mean if a **human** is an "**outlier**"?



Bansal, Ankan, et al. "Zero-shot object detection." *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.

16

Healthy Machine Learning in Health

what **models** are **healthy**?

what **healthcare** is **healthy**?

what **behaviors** are **healthy**?

Creating actionable insights in human health.

Bias Is Part of the Clinical Landscape

- How does/should ML interact with fairness/health[1,2,3,4,5]?

This Issue | Views 12,435 | Citations 41 | Altmetric 174

Viewpoint

August 11, 2015

Racial Bias in Health Care and Health
Challenges and Opportunities

David R. Williams, PhD, MPH[1,2]; Ronald Wyatt, MD, MHA[3]

Author Affiliations

JAMA. 2015;314(6):555-556. doi:10.1001/jama.2015.9260

J Palliat Med. 2013 Nov; 16(11): 1329–1334.
doi: 10.1089/jpm.2013.9468

PMCID: PMC3822363
PMID: 24073685

Racial and Ethnic Disparities in Palliative Care

Kimberly S. Johnson, MD, MHS[1,2]

Author information ▶ Article notes ▶ Copyright and License information ▶ Disclaimer

This article has been cited by other articles in PMC.

2

The Girl Who Cried Pain: A Bias Against Women in the Treatment of Pain

Diane E. Hoffmann and Anita J. Tarzian

Am J Public Health. 2007 February; 97(2): 247–251.
doi: 10.2105/AJPH.2005.072975

PMCID: PMC1781382
PMID: 17194867

The Black–White Disparity in Pregnancy-Related Mortality From 5 Conditions: Differences in Prevalence and Case-Fatality Rates

Myra J. Tucker, BSN, MPH, Cynthia J. Berg, MD, MPH, William M. Callaghan, MD, MPH, and Jason Hsia, PhD

Author information ▶ Article notes ▶ Copyright and License information ▶ Disclaimer

Obes Rev. 2015 Apr;16(4):319-26. doi: 10.1111/obr.12266. Epub 2015 Mar 5.

Impact of weight bias and stigma on quality of care and outcomes for patients with obesity.

Phelan SM[1], Burgess DJ, Yeazel MW, Hellerstedt WL, Griffin JM, van Ryn M.

⊕ Author information

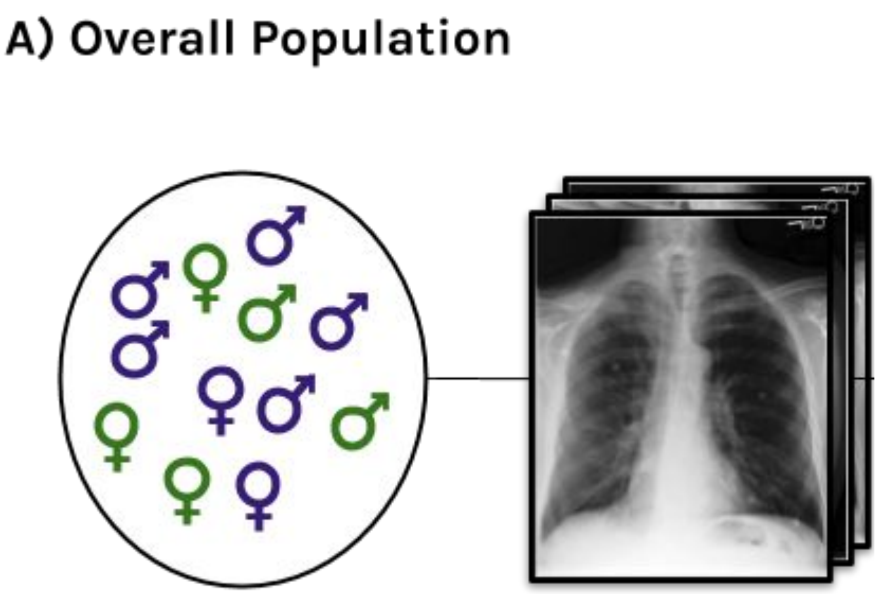
[1] Continuous State-Space Models for Optimal Sepsis Treatment - Deep Reinforcement Learning … (MLHC/JMLR 2017);
[2] Modeling Mistrust in End-of-Life Care (MLHC 2018/FATML 2018 Workshop);
[3] The Disparate Impacts of Medical and Mental Health with AI. (AMA Journal of Ethics 2019);
[4] ClinicalVis Project with Google Brain. (*In submission);

imes

MIT EECS CSAIL

18

Model-based Chest X-Ray Diagnosis

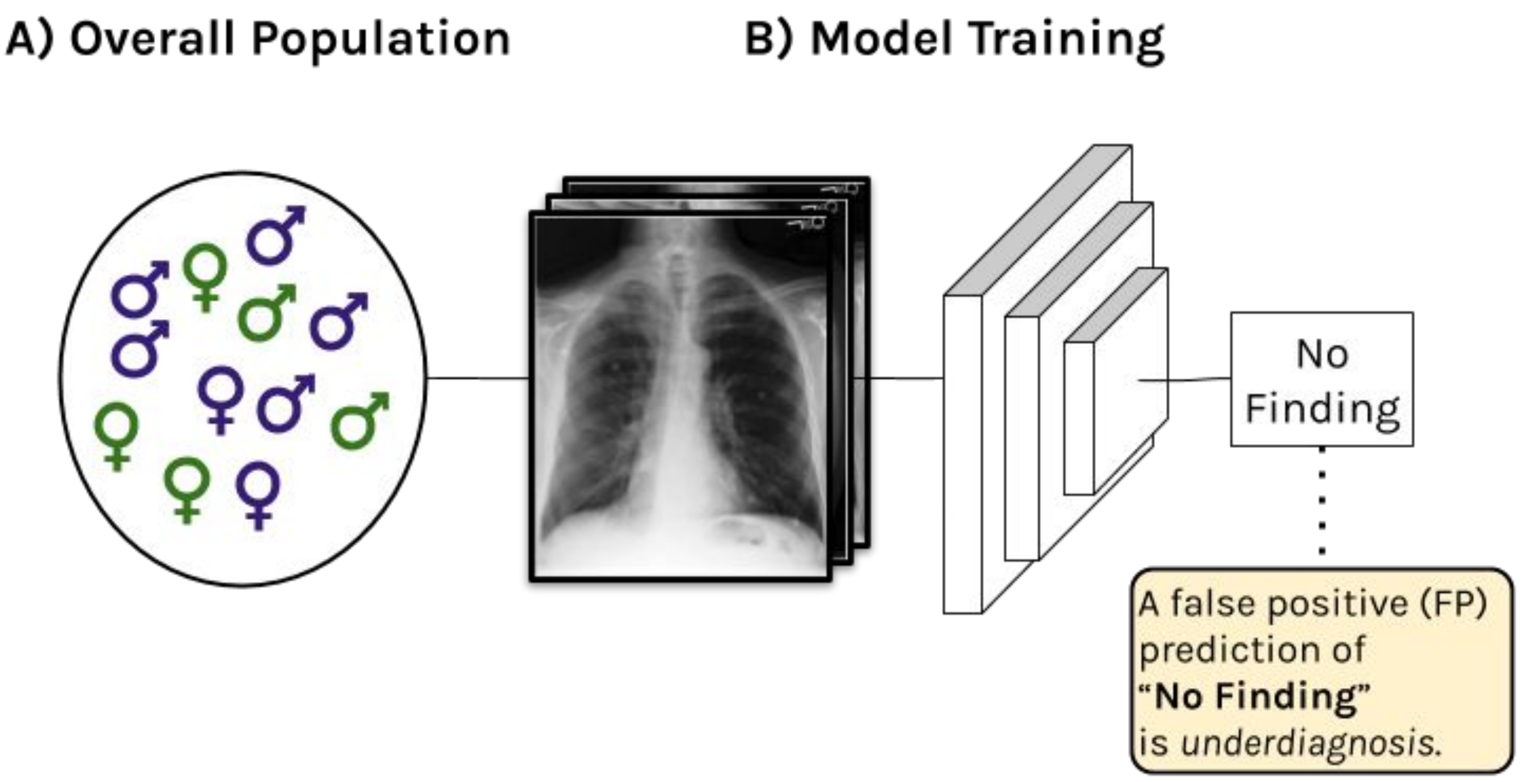


- Take 3 large **chest x-ray** datasets (707,626 images).

[1] Seyyed-Kalantari, Zhang, Liu, McDermott, Chen, Ghassemi. “Medical imaging algorithms exacerbate biases in underdiagnosis.” Nature Medicine 2021. To appear.

19

Model-based Chest X-Ray Diagnosis

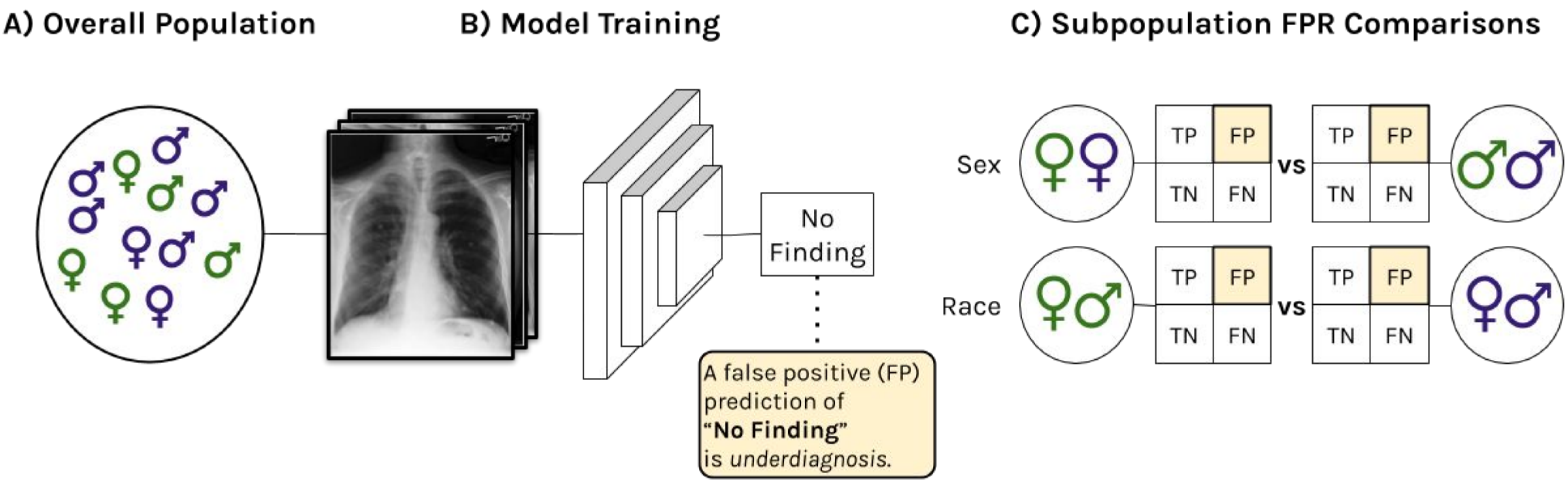


- Take 3 large **chest x-ray** datasets (707,626 images).
- Train a DenseNet to predict a "**No Finding**" label, e.g., model says patient is healthy.

[1] Seyyed-Kalantari, Zhang, Liu, McDermott, Chen, Ghassemi. "Medical imaging algorithms exacerbate biases in underdiagnosis." Nature Medicine 2021. To appear.

20

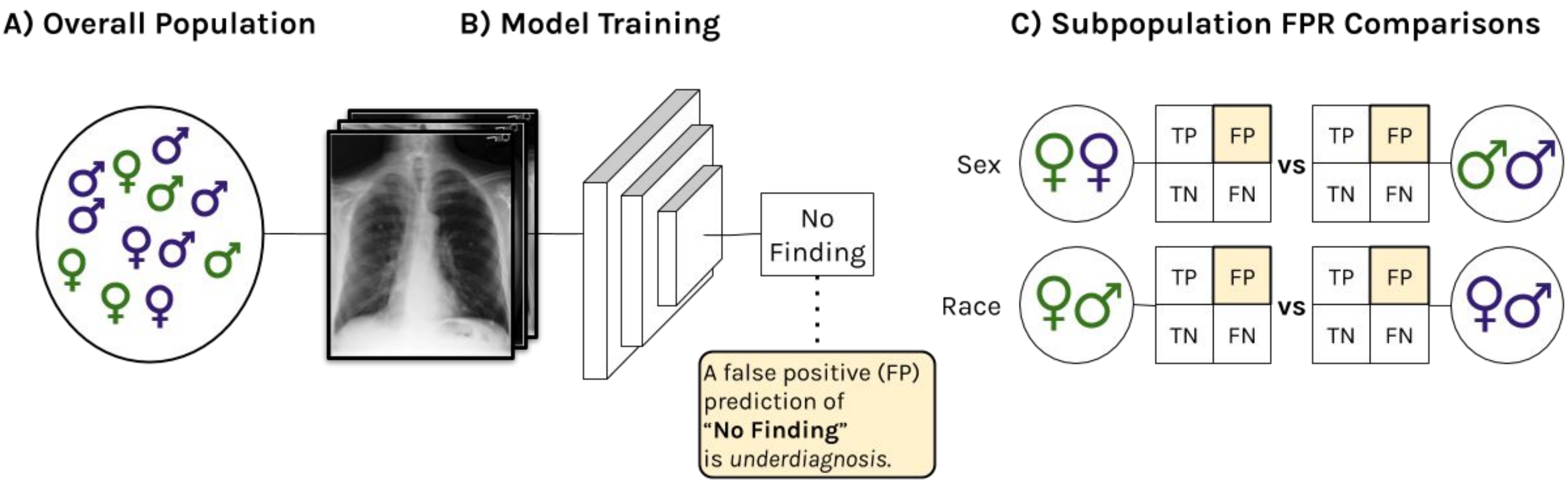
Model-based Chest X-Ray Diagnosis



- Take 3 large **chest x-ray** datasets (707,626 images).
- Train a DenseNet to predict a "**No Finding**" label, e.g., model says patient is healthy.
- Compare false positive rate (FPR) in different subpopulations to examine model **underdiagnosis rates**.

[1] Seyyed-Kalantari, Zhang, Liu, McDermott, Chen, Ghassemi. "Medical imaging algorithms exacerbate biases in underdiagnosis." Nature Medicine 2021. To appear.

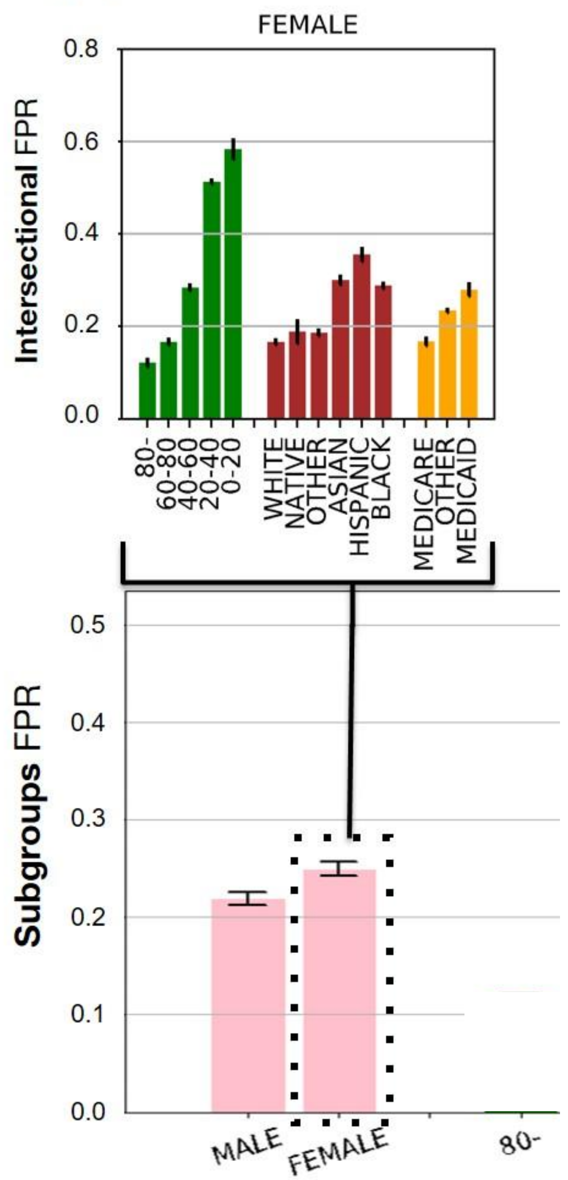
Model-based Chest X-Ray Diagnosis



Higher model underdiagnosis rates on one **subpopulation**, such as **female patients**, would lead to a **higher rate** of **no treatment** for those patients if the model were **deployed**.

[1] Seyyed-Kalantari, Zhang, Liu, McDermott, Chen, Ghassemi. "Medical imaging algorithms exacerbate biases in underdiagnosis." Nature Medicine 2021. To appear.

Automating CheXclusion With EHR + ML

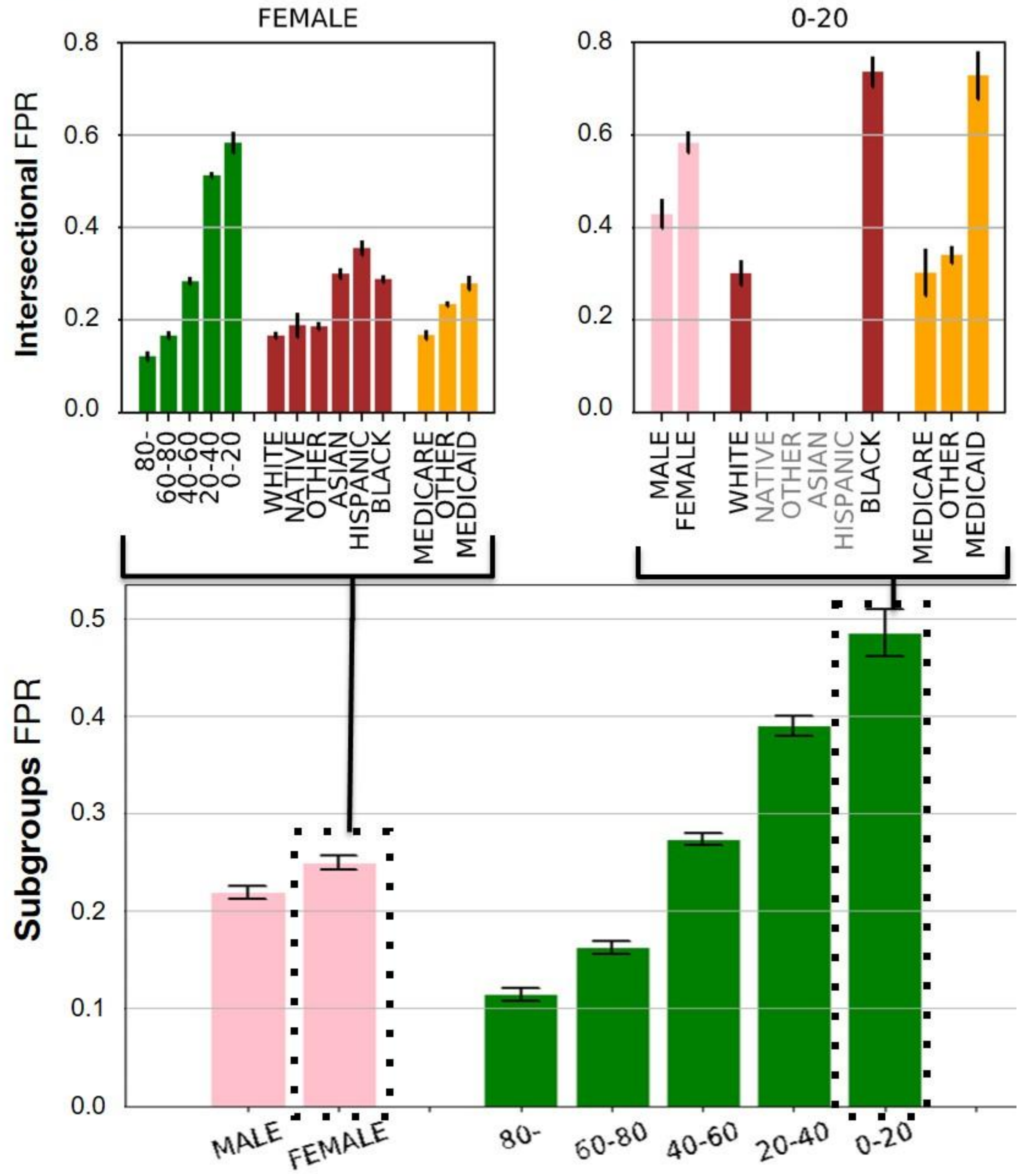


- Largest underdiagnosis rates in Female

[1] Seyyed-Kalantari, Zhang, Liu, McDermott, Chen, Ghassemi. "Medical imaging algorithms exacerbate biases in underdiagnosis." Nature Medicine 2021. To appear.

23

Automating CheXclusion With EHR + ML

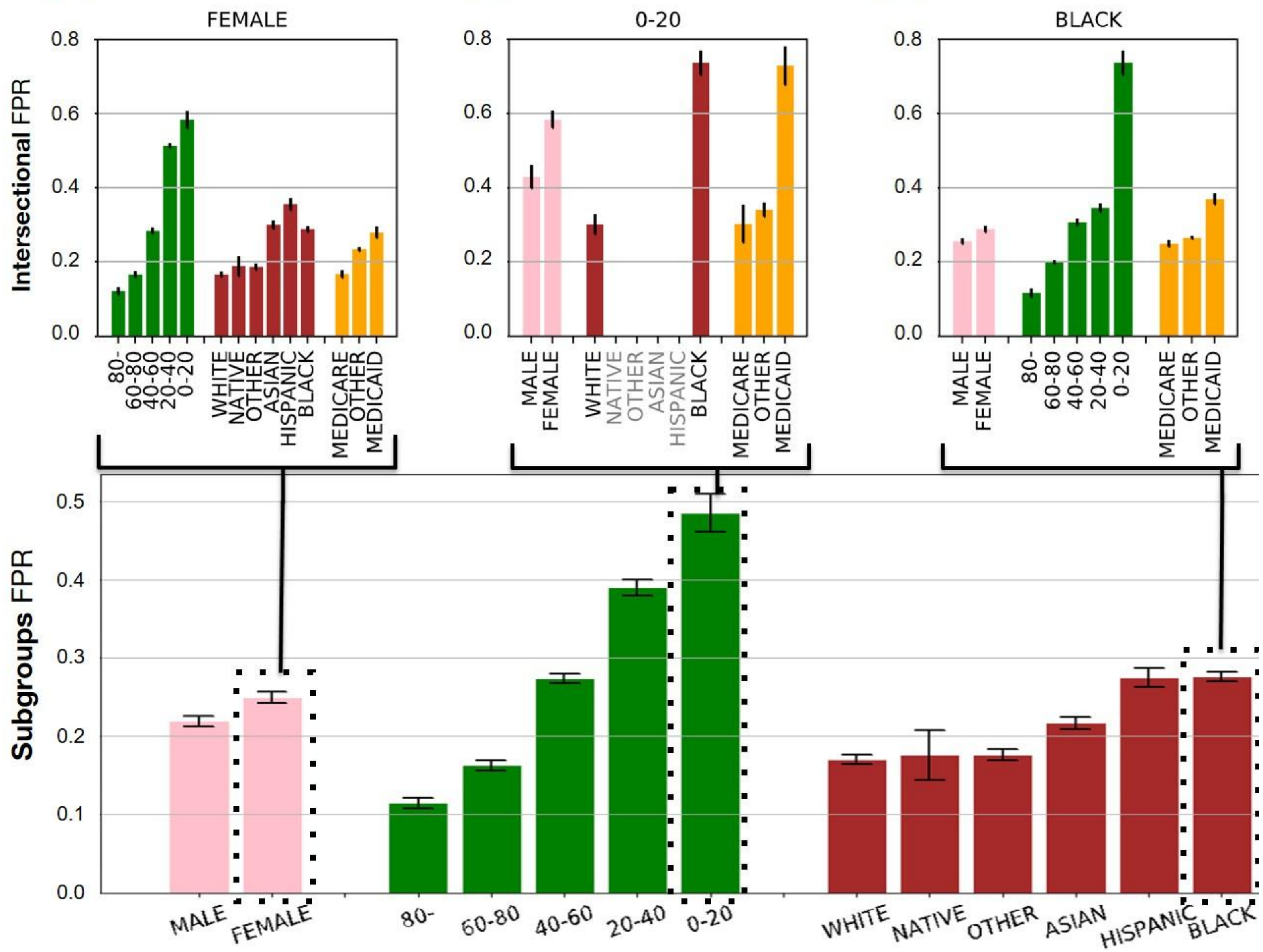


- Largest underdiagnosis rates in Female, 0-20

[1] Seyyed-Kalantari, Zhang, Liu, McDermott, Chen, Ghassemi. "Medical imaging algorithms exacerbate biases in underdiagnosis." Nature Medicine 2021. To appear.

24

Automating CheXclusion With EHR + ML

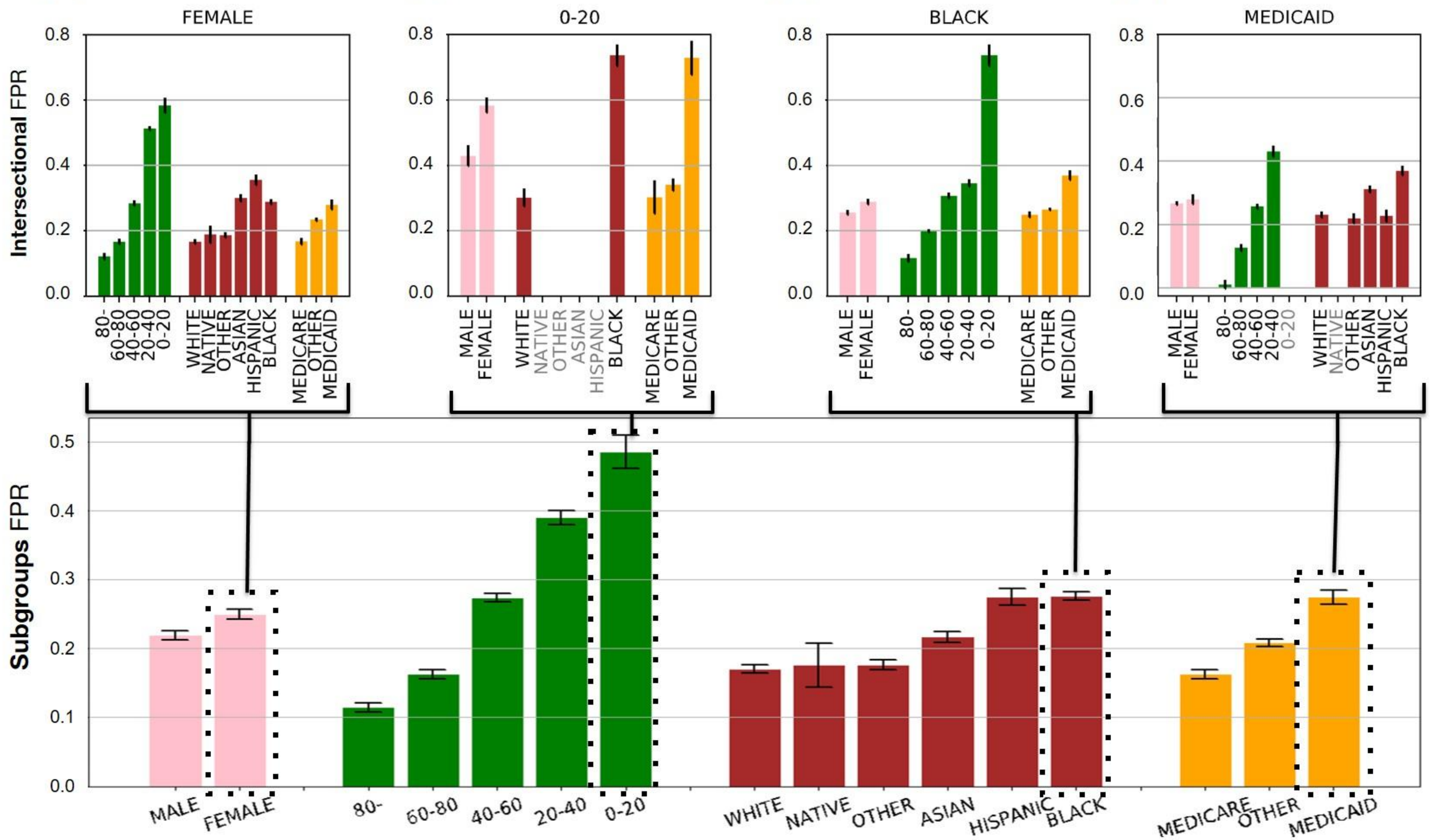


- Largest underdiagnosis rates in Female, 0-20, Black

[1] Seyyed-Kalantari, Zhang, Liu, McDermott, Chen, Ghassemi. “Medical imaging algorithms exacerbate biases in underdiagnosis.” Nature Medicine 2021. To appear.

25

Automating CheXclusion With EHR + ML



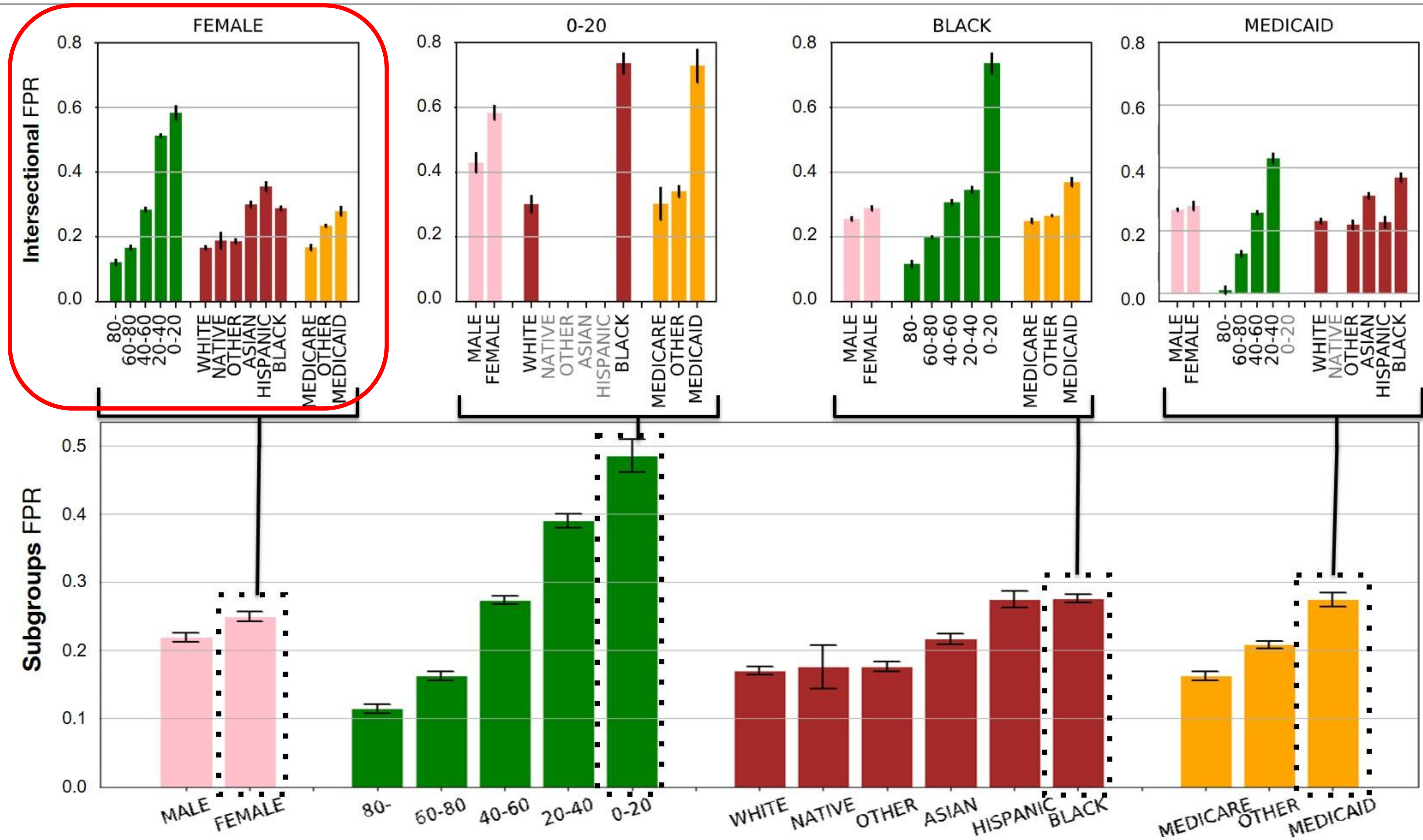
- Largest underdiagnosis rates in Female, 0-20, Black, and Medicaid insurance patients.

[1] Seyyed-Kalantari, Zhang, Liu, McDermott, Chen, Ghassemi. "Medical imaging algorithms exacerbate biases in underdiagnosis." Nature Medicine 2021. To appear.

imes MIT EECS CSAIL

26

Automating CheXclusion With EHR + ML

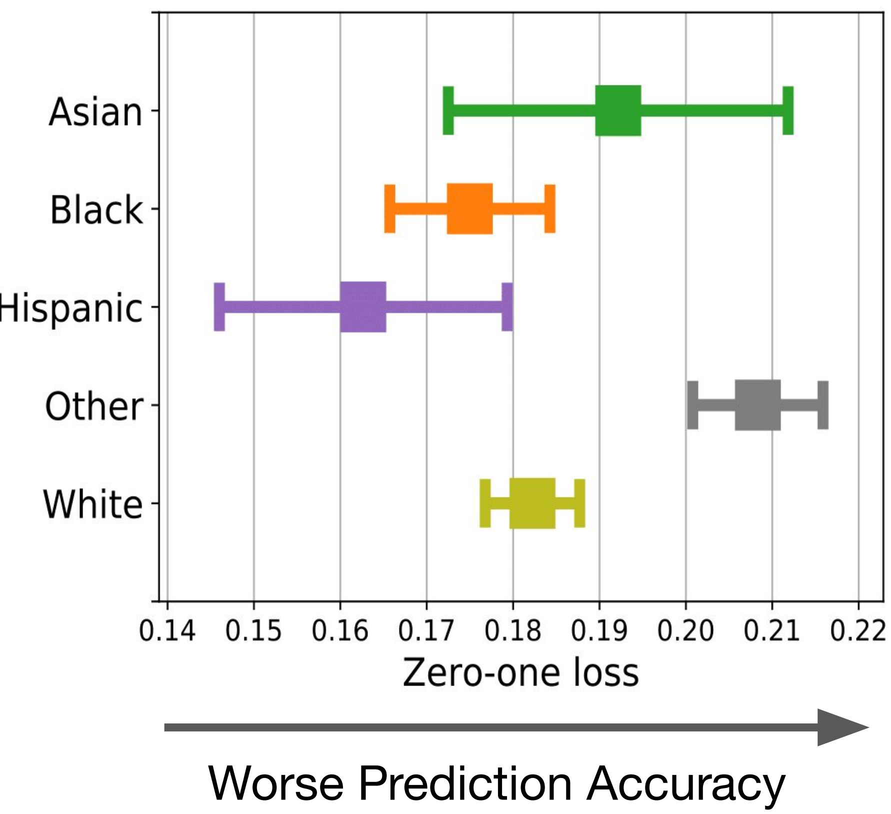


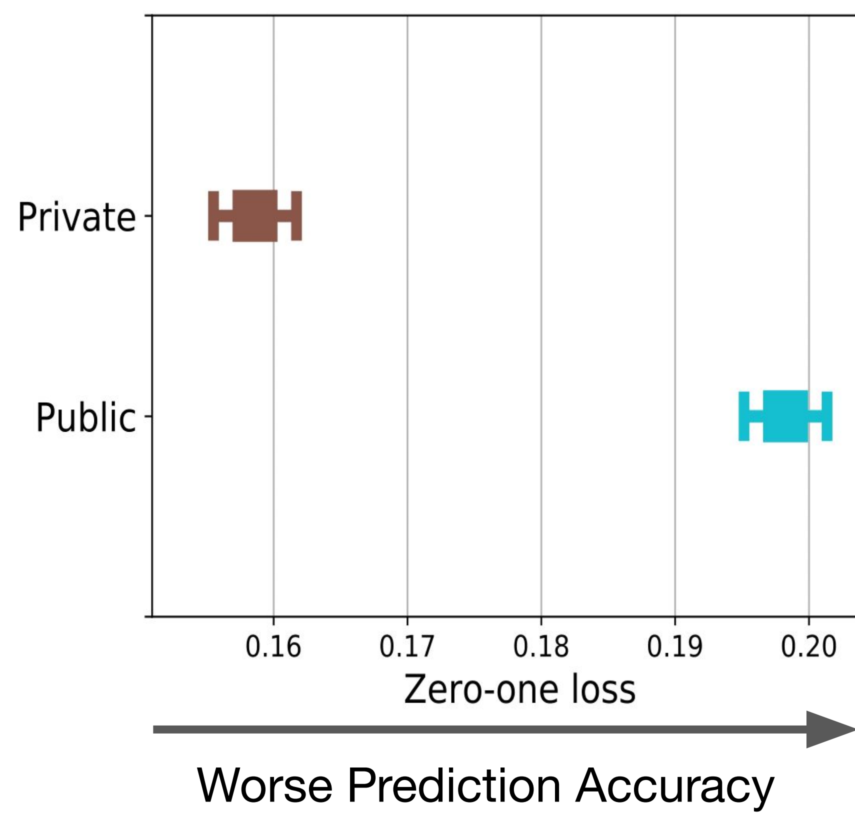
- **Intersectional** identities are often underdiagnosed even more heavily than the aggregate group, e.g., **Black or Hispanic female patients** are **underdiagnosed more** than White female patients.

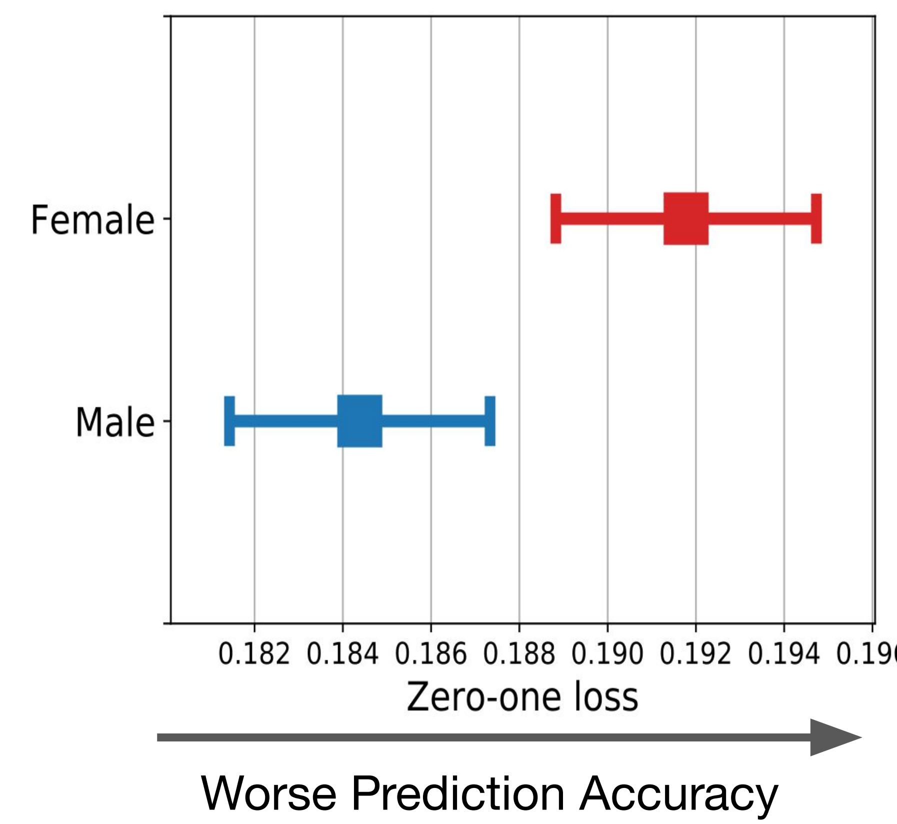
[1] Seyyed-Kalantari, Zhang, Liu, McDermott, Chen, Ghassemi. "Medical imaging algorithms exacerbate biases in underdiagnosis." Nature Medicine 2021. To appear.

Auditing Fairness In Predictive Models

- Significant differences in model accuracy for race, sex, and insurance type in **ICU notes** and insurance type in **psychiatric notes**.







[1] Chen, Szolovits, Ghassemi. "Can AI Help Reduce Disparities in General Medical and Mental Health Care?." *AMA journal of ethics* 21.2 (2019): 167-179.

28

Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings

Prompt: `[**RACE**] pt became belligerent and violent . sent to [**TOKEN**] [**TOKEN**]`

[1] Zhang, Lu, Abdallah, Ghassemi. "Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings". ACM CHIL 2020.

29

Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings

Prompt: **[**RACE**] pt became belligerent and violent . sent to [**TOKEN**] [**TOKEN**]**

SciBERT: caucasian pt became belligerent and violent . sent to **hospital .**
white pt became belligerent and violent . sent to **hospital .**

[1] Zhang, Lu, Abdallah, Ghassemi. "Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings". ACM CHIL 2020.

30

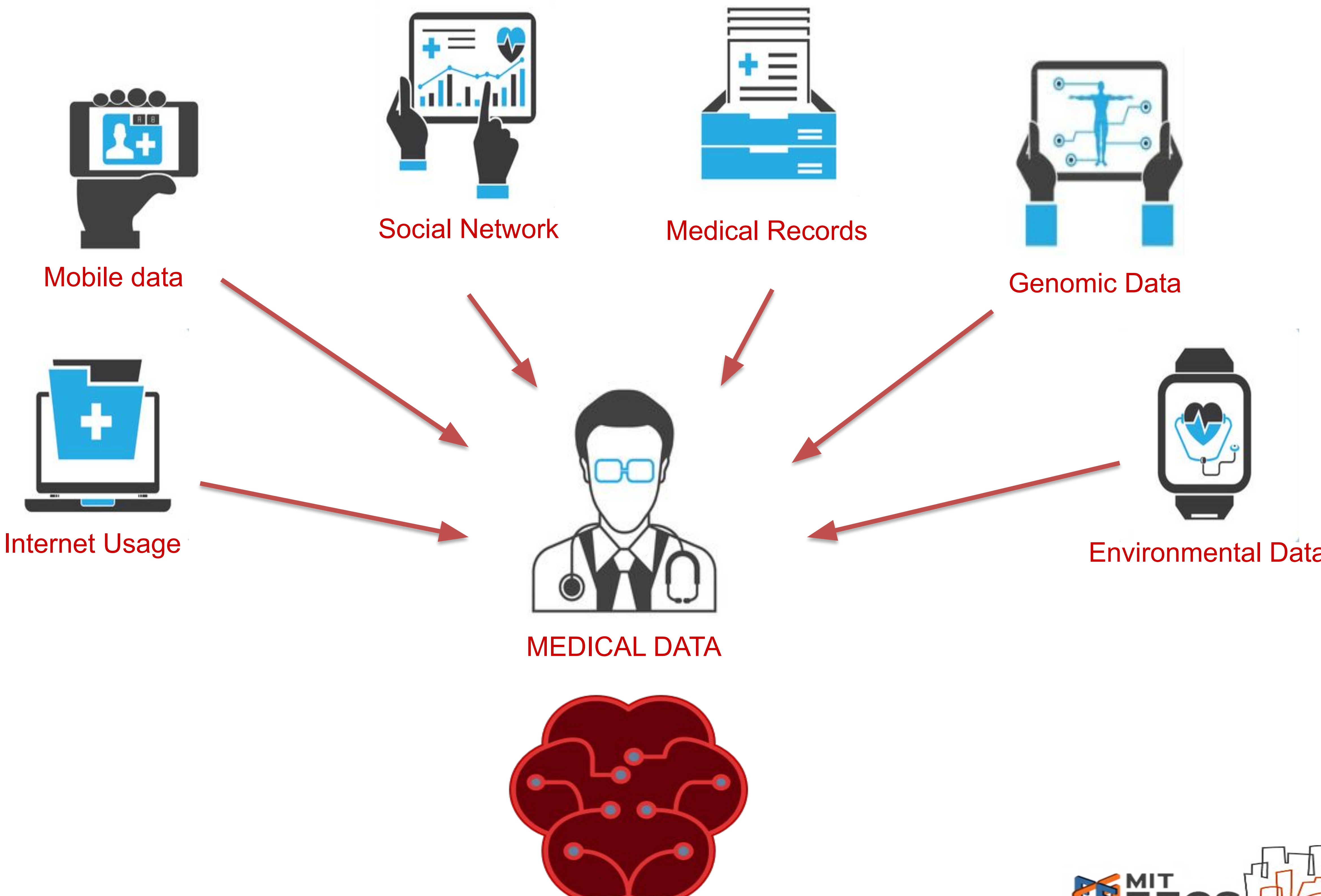
Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings

Prompt: **[**RACE**] pt became belligerent and violent . sent to [**TOKEN**] [**TOKEN**]**

SciBERT:
caucasian pt became belligerent and violent . sent to **hospital .**
white pt became belligerent and violent . sent to **hospital .**
african pt became belligerent and violent . sent to **prison .**
african american pt became belligerent and violent . sent to **prison .**
black pt became belligerent and violent . sent to **prison .**

[1] Zhang, Lu, Abdallah, Ghassemi. "Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings". ACM CHIL 2020.

Complex **Health** Generates Complex Data



Healthy Machine Learning in Health

what **models** are **healthy**?

what **healthcare** is **healthy**?

what **behaviors** are **healthy**?

Creating actionable insights in human health.

33

How Much Behavioral Variation Could There Be?

- eBay auction study looked at iPods where researchers randomly **varied the skin color on the hand** holding the iPod.

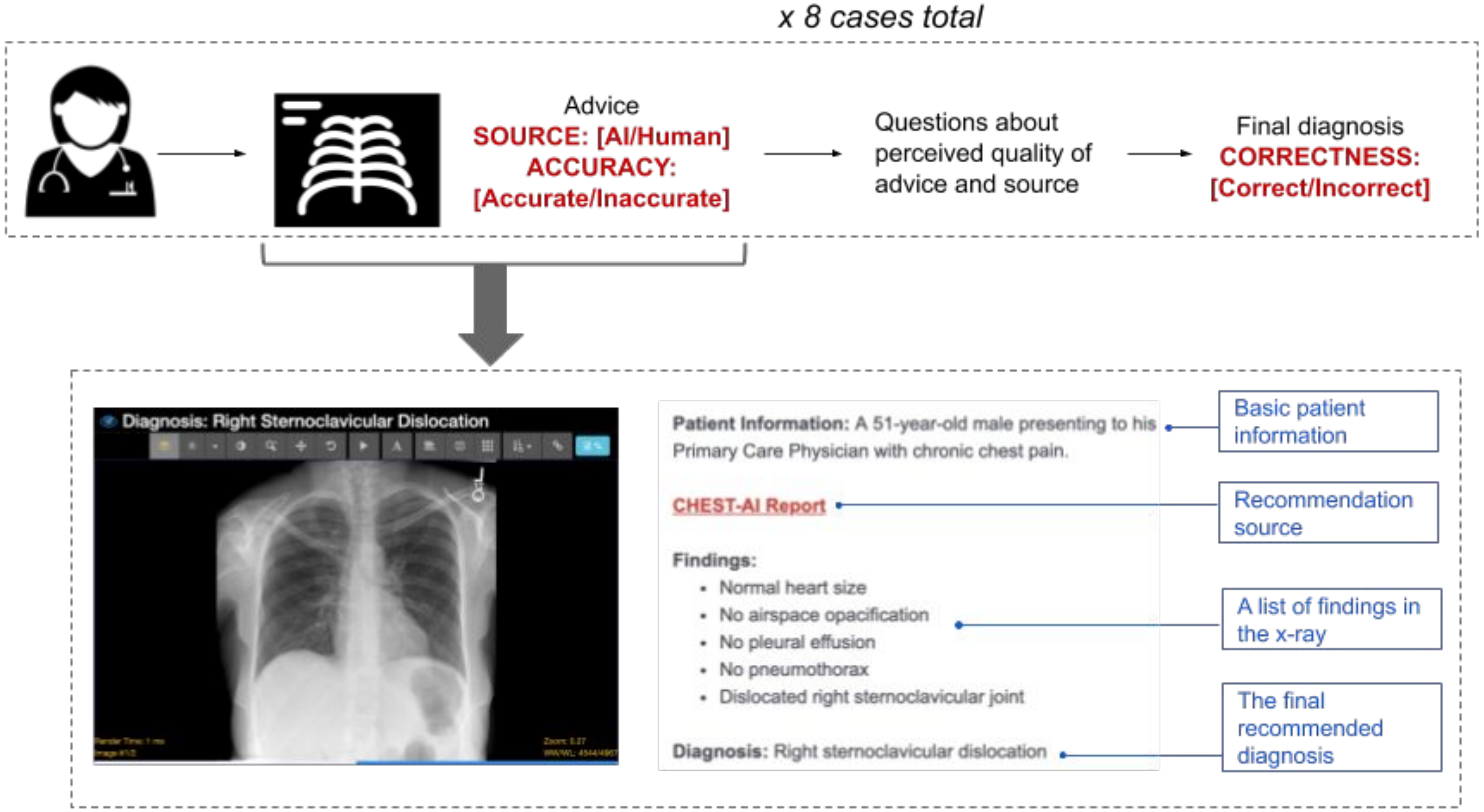
- A white hand holding the iPod received 21 percent more offers than a black hand.

Doleac, Jennifer L., and Luke CD Stein. "The visible hand: Race and online market outcomes." The Economic Journal 123.572 (2013): F469-F492.

34

Diagnostic X-Ray Advice In Expert/Non-Experts

- What is clinical interaction with “AI” vs. “human” advice?



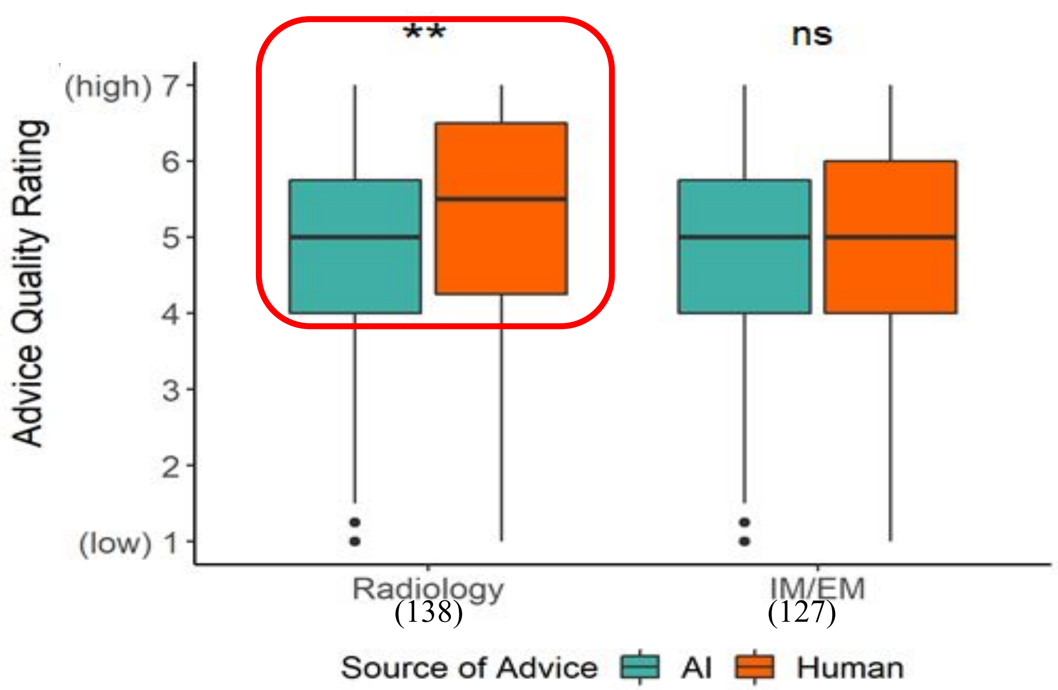
- Evaluate expert (radiologists) vs. non-expert (internal/emergency medicine) clinicians.

Gaube, Susanne, et al. "Do as AI say: susceptibility in deployment of clinical decision-aids." NPJ digital medicine 4.1 (2021): 1-8.

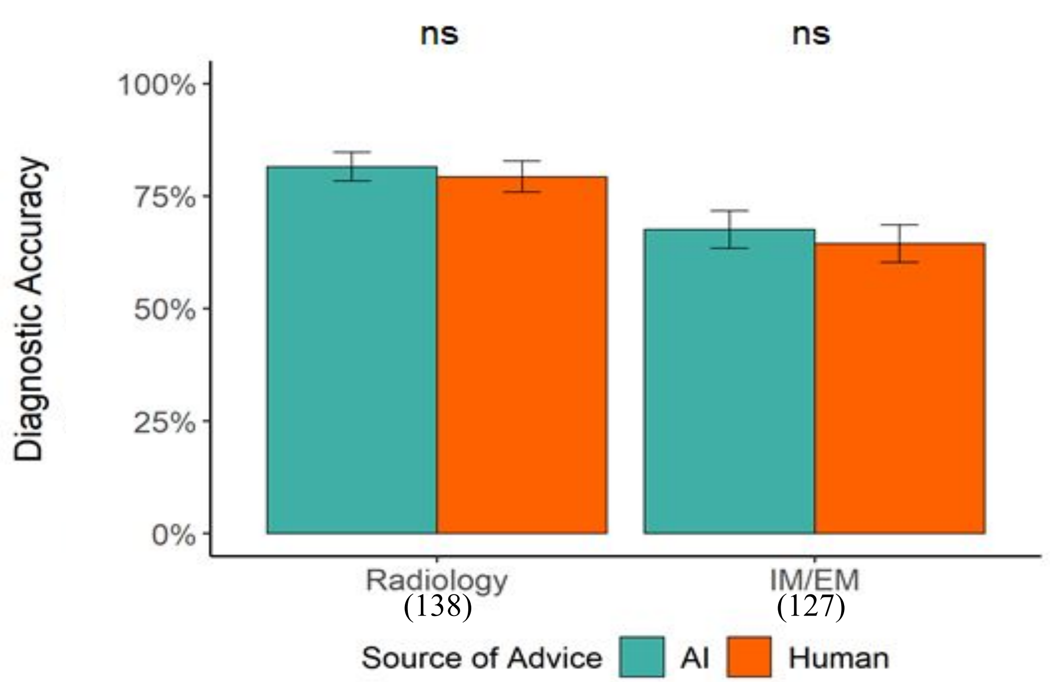
Expertise and Algorithmic **Aversion**

- Task-expert radiologists **rate "AI" advice lower** than "human" advice.
- Physicians across expertise levels **failed to dismiss** incorrect advice.

Expert Rate AI Worse



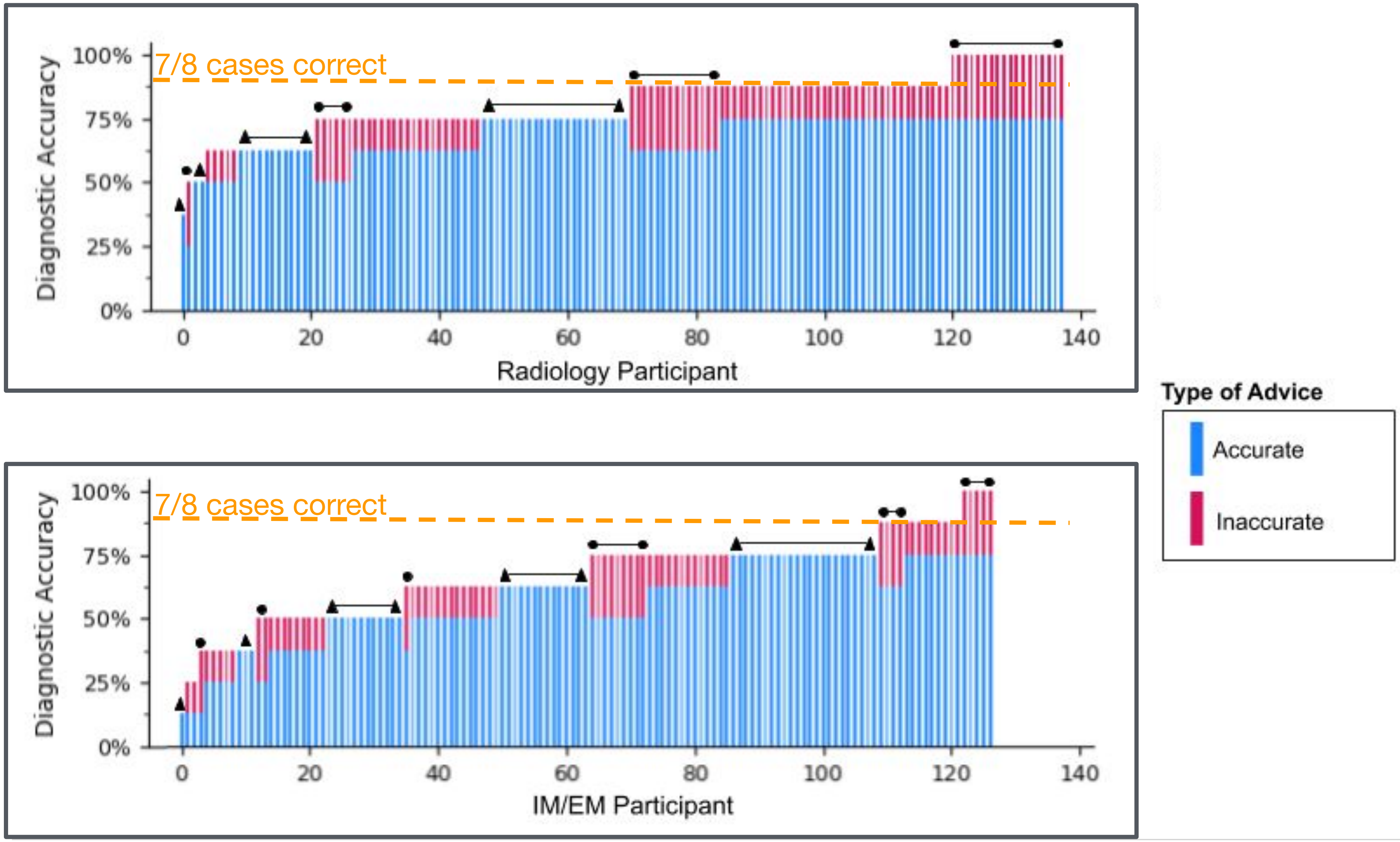
Similar Accuracy



Gaube, Susanne, et al. "Do as AI say: susceptibility in deployment of clinical decision-aids." NPJ digital medicine 4.1 (2021): 1-8.

Expertise and **Susceptibility**

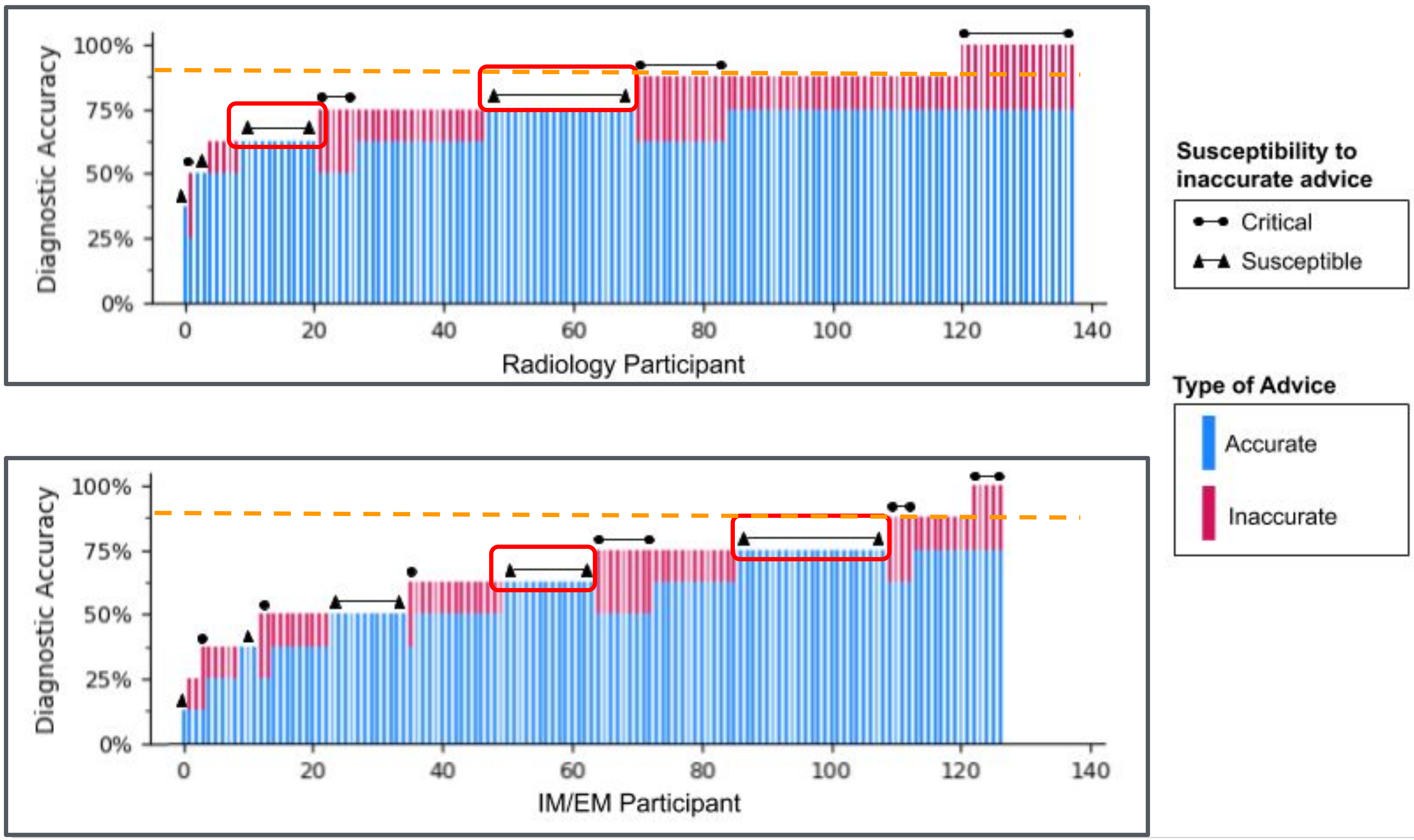
- Experts have better **diagnostic accuracy**; ~½ get ⅞ cases correct.



Gaube, Susanne, et al. "Do as AI say: susceptibility in deployment of clinical decision-aids." NPJ digital medicine 4.1 (2021): 1-8.

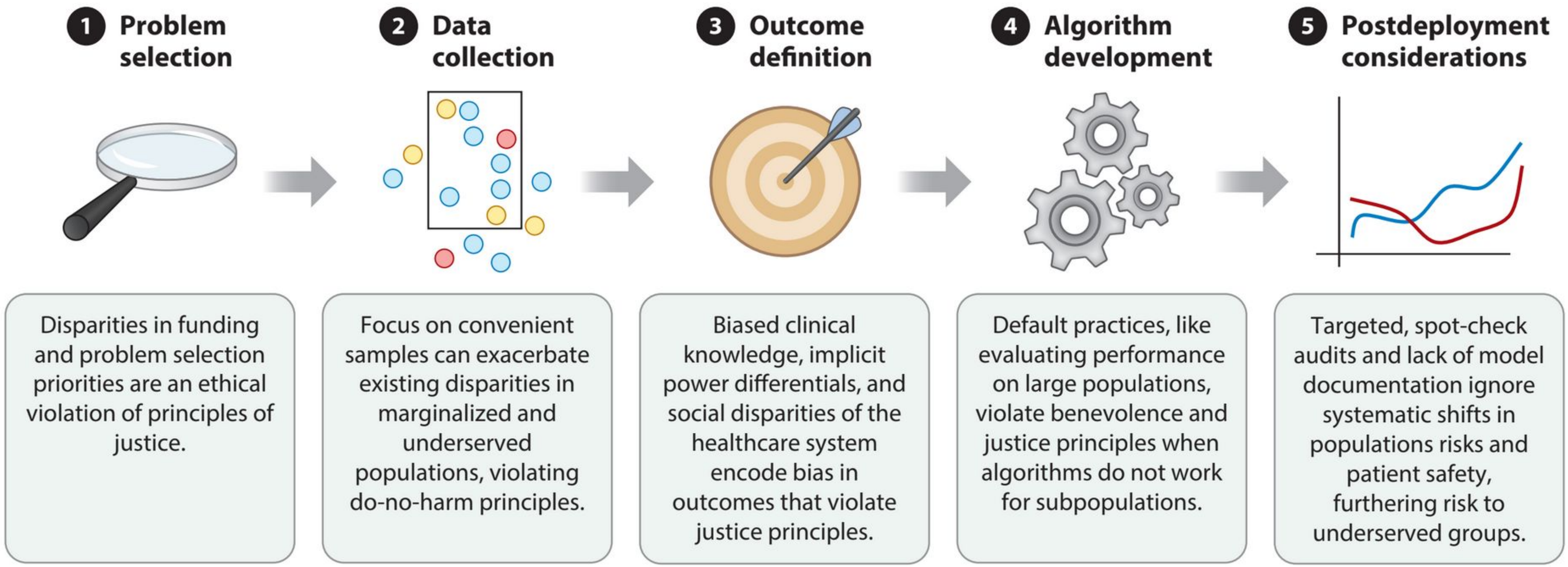
Expertise and **Susceptibility**

- Experts have better **diagnostic accuracy**; ~½ get ⅞ cases correct.
- Some doctors are more **susceptible to incorrect advice** than others.



Gaube, Susanne, et al. "Do as AI say: susceptibility in deployment of clinical decision-aids." NPJ digital medicine 4.1 (2021): 1-8.

No Simple Fixes for Ethical ML in Health

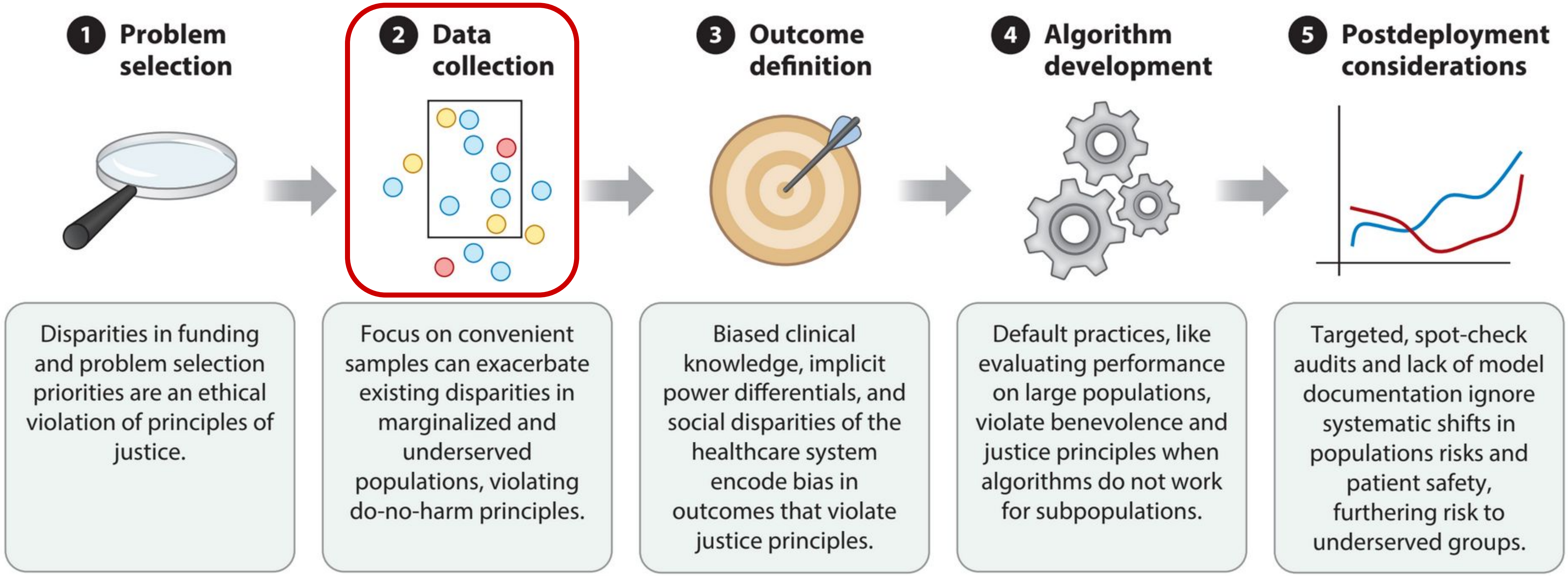


- This is an **on-going** process that requires improvement.
- Requires engagement on many levels by diverse teams.

Chen, Irene Y., et al. "Ethical Machine Learning in Healthcare." *Annual Review of Biomedical Data Science* 4 (2020).

39

No Simple Fixes for Ethical ML in Health

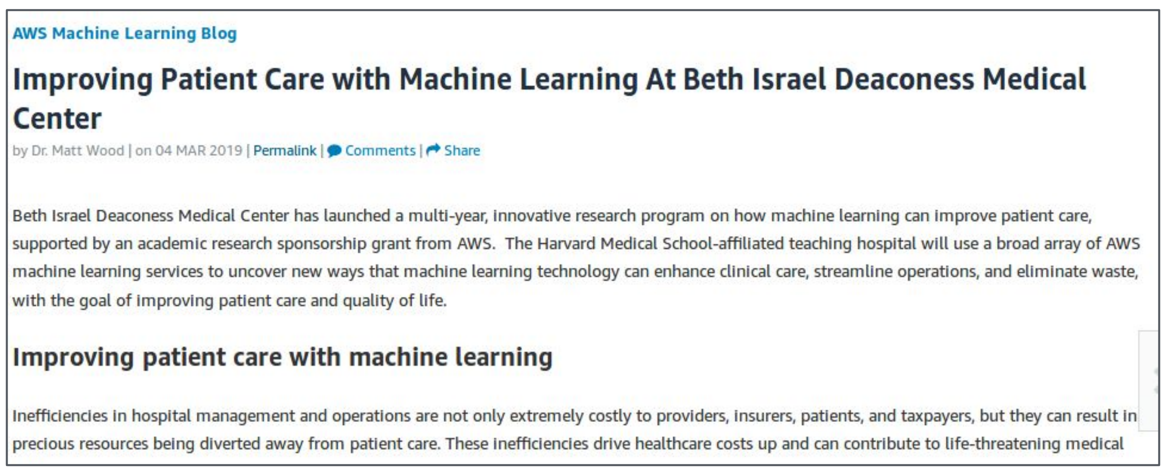


- This is an **on-going** process that requires improvement.
- Requires engagement on many levels by diverse teams.

Chen, Irene Y., et al. "Ethical Machine Learning in Healthcare." *Annual Review of Biomedical Data Science* 4 (2020).

Availability of Em**bodied** Data

- All data is valuable; embodied health data particularly so.
- Robust, private, fair, high-quality algorithms require large-scale diverse datasets for research use.


AWS Machine Learning Blog

Improving Patient Care with Machine Learning At Beth Israel Deaconess Medical Center

by Dr. Matt Wood | on 04 MAR 2019 | Permalink | Comments | Share

Beth Israel Deaconess Medical Center has launched a multi-year, innovative research program on how machine learning can improve patient care, supported by an academic research sponsorship grant from AWS. The Harvard Medical School-affiliated teaching hospital will use a broad array of AWS machine learning services to uncover new ways that machine learning technology can enhance clinical care, streamline operations, and eliminate waste, with the goal of improving patient care and quality of life.

Improving patient care with machine learning

Inefficiencies in hospital management and operations are not only extremely costly to providers, insurers, patients, and taxpayers, but they can result in precious resources being diverted away from patient care. These inefficiencies drive healthcare costs up and can contribute to life-threatening medical


Amazon Comprehend Medical

Extract information from unstructured medical text accurately and quickly

No machine learning experience required

Get started with Amazon Comprehend Medical

Amazon Comprehend Medical is a natural language processing service that makes it easy to use machine learning to extract relevant medical information from unstructured text. Using Amazon Comprehend Medical, you can quickly and accurately gather information, such as medical condition, medication, dosage, strength, and frequency from a variety of sources like doctors' notes, clinical trial reports, and patient health records.

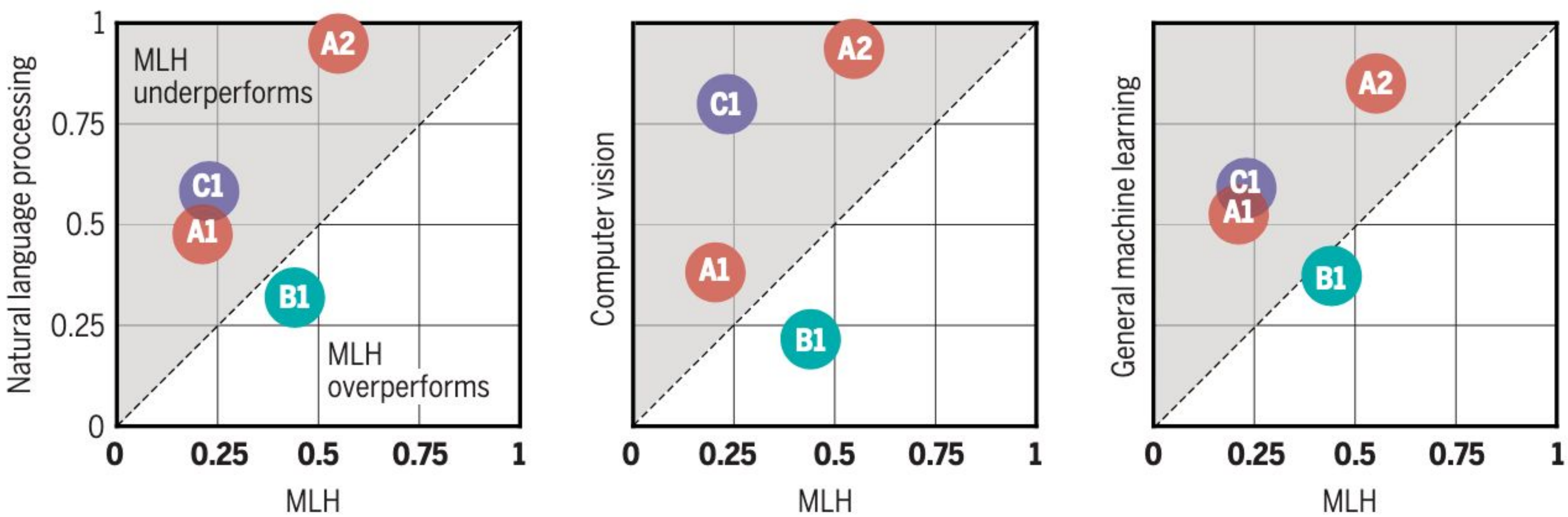

Google Tries to Patent Healthcare Deep Learning, EHR Analytics

Google has applied for a sweeping patent including the fundamentals of deep learning and EHR analytics in the healthcare industry.

Source: Google

41

Health Lags Other ML Subfields in Reproducibility



- ML in Health lags in reproducibility metrics:
  - Releasing code (A1)
  - Releasing data (A2)
  - Leveraging multiple data-sets (C1)

Evaluation metrics

A Technical reproducibility
1 Code available
2 Public dataset

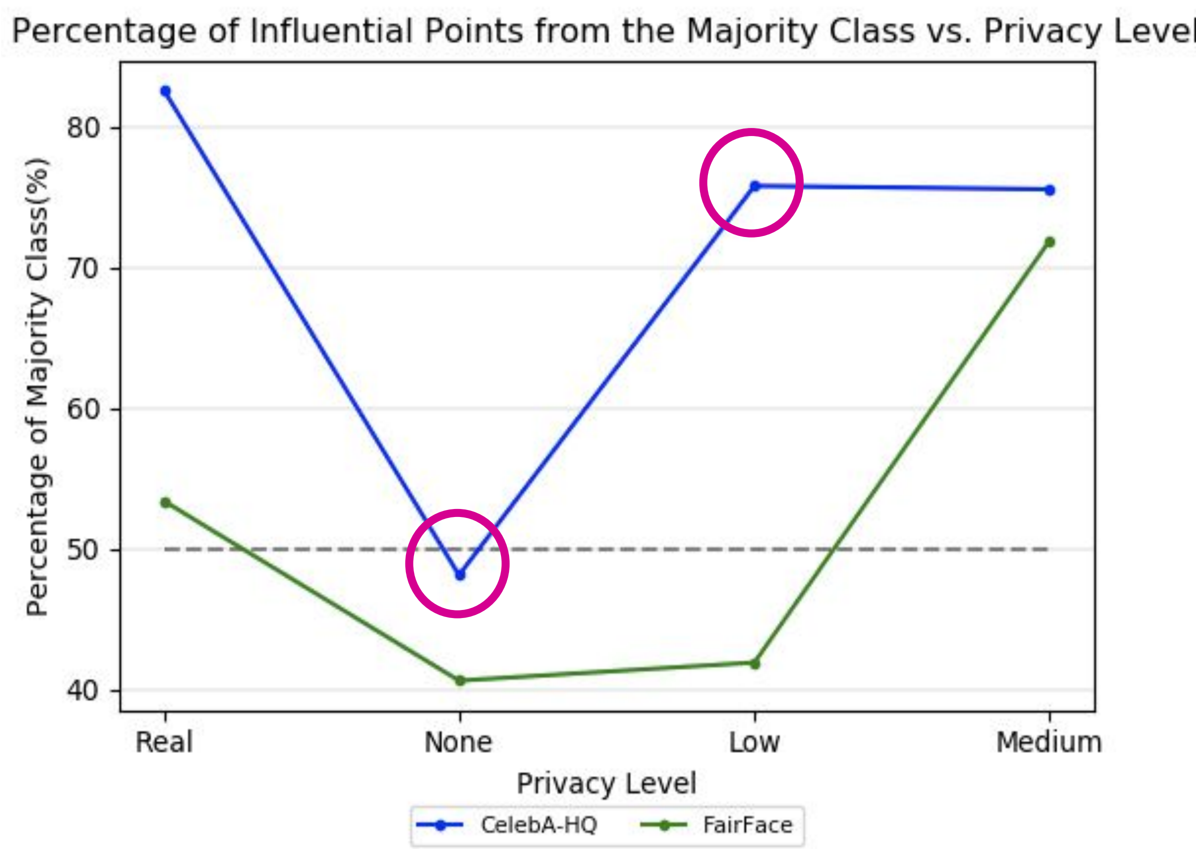
B Statistical reproducibility
1 Variance reported

C Conceptual reproducibility (replicability)
1 Multiple datasets

McDermott, Matthew BA, et al. "Reproducibility in machine learning for health research: Still a ways to go." Science Translational Medicine 13.586 (2021).

Synthetic Data Is Not a Robust Solution

- Biased datasets can have **disparate impacts** on minority downstream classification influence...
  - Even when the real dataset is not directly used
  - Even when the synthetic dataset used for the training is balanced

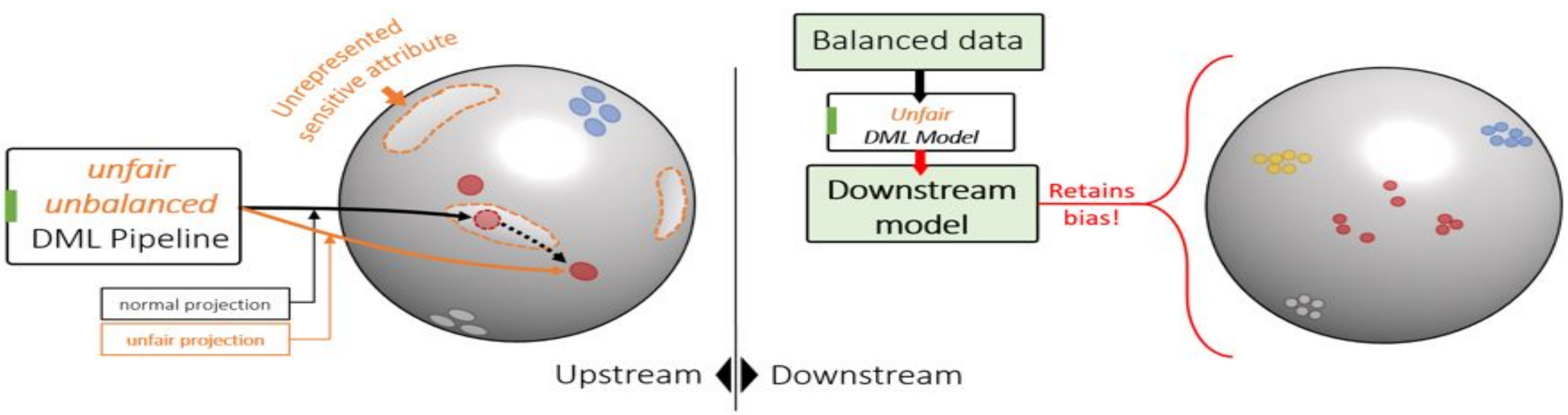


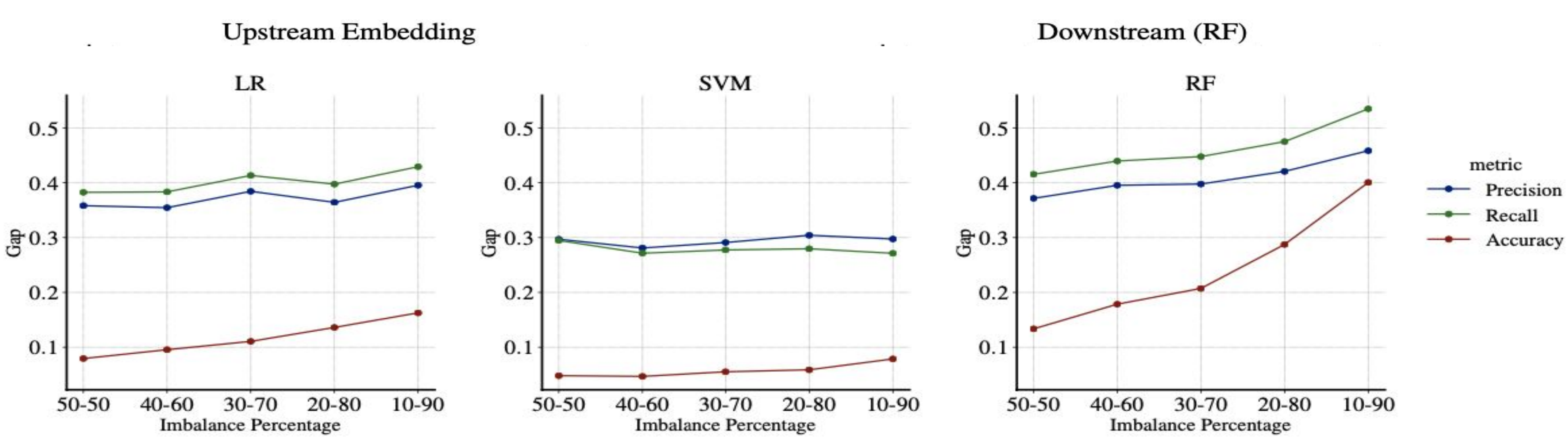
- Supplementing or **replacing datasets** with **synthetic data** does not mitigate the **fairness concerns** caused by the existing biases in imbalanced datasets.

Cheng, Victoria, et al. "Can You Fake It Until You Make It? Impacts of Differentially Private Synthetic Data on Downstream Classification Fairness." *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 2021.

Neither Is Post-Hoc Balancing

- Bias in data causes **asymmetric** upstream **embeddings**.
- Biased embeddings **impact** downstream tasks, even with **rebalancing**!

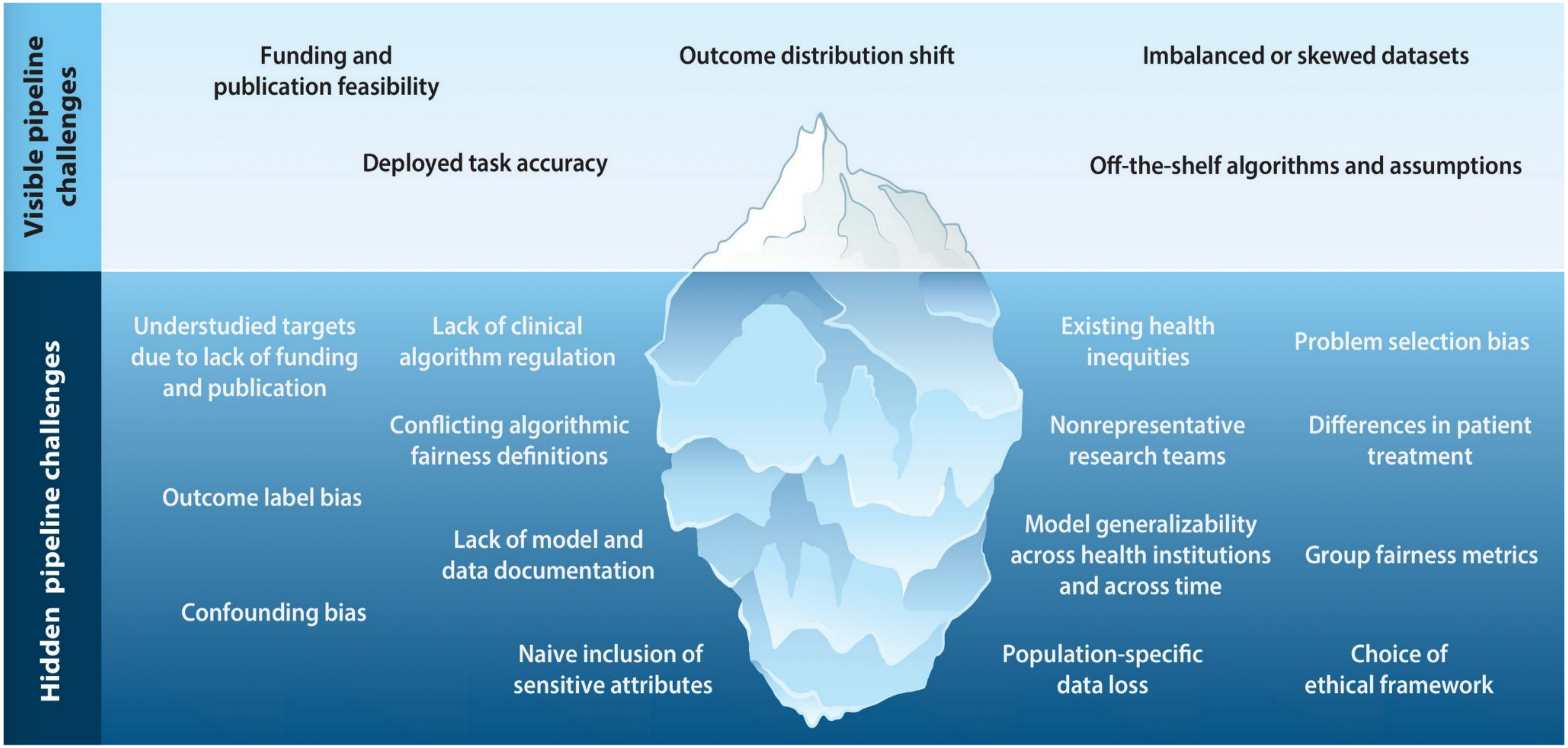




Dullerud, Natalie, et al. "Is Fairness Only Metric Deep?" *In Submission.*

The Tip of The AI-ceberg

- Some issues lie very far under the waterline, and require introspection.



Chen, Irene Y., et al. "Ethical Machine Learning in Healthcare." *Annual Review of Biomedical Data Science* 4 (2020).

Dr.
Marzyeh
Ghassemi

Healthy ML @ MIT IMES EECS.CSAIL

Students


Natalie Dullerud


Vinith Suriyakumar


Haoran Zhang

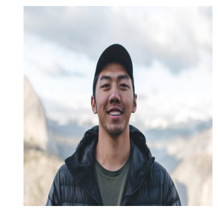

Aparna Balagopalan

Kimia Hamideh


Sindhu Gowda


Minfan Zhang


Taylor Killian


Nathan Ng


Bret Nestor

Collaborators (Technical and Clinical)


Anna Goldenburg


Mehdi Fatemi


Shalmali Joshi


Amol Verma


Fahad Razak


Muhammad Mamdani


Leo Celi

Funding Sources

- CIFAR AI Chair & CIFAR Azrieli Global Scholar
- Microsoft Research
- Helmsley Trust
- Wellcome Trust
- Quanta Computing

Healthy Machine Learning in Health

what **models** are **healthy**?

what **healthcare** is **healthy**?

what **behaviors** are **healthy**?

Creating actionable insights in human health.

47