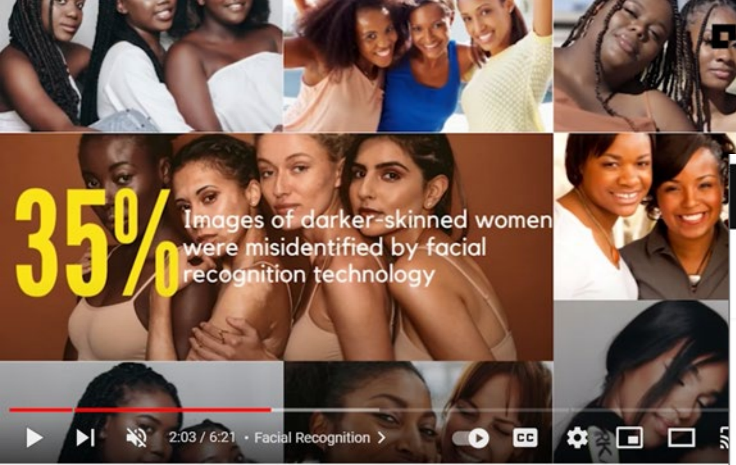
Recognizing and Integrating Social Good into the AI Development Lifecycle

November 1, 2022

Bradley Malin, Ph.D.
Accenture Professor of Biomedical Informatics, Biostatistics, and Computer Science
Co-Director, Health Data Science Center
Vanderbilt University

MPI: Center for Genetic Privacy and Identity in Community Settings (NIH CEER)
AIM-AHEAD Infrastructure Core (NIH)
Bridge2AI Ethics and Trustworthy AI Core (NIH)



Facial Recognition Is Sexist, Racist And Biased

584 views... 18 DISLIKE SHARE DOWNLOAD CLIP

Analytics India Magazine

The Washington Post
Democracy Dies in Darkness

Tech Help Desk Future of Transportation Innovations Internet Culture Space Tech Policy Video Gaming

INNOVATIONS

These robots we[...] became racist an[...]

As billions flow into robotics, researchers who c[...] society

By Pranshu Verma
July 16, 2022 at 6:00 a.m. EDT

2022

Artificial vs. human intelligence concept. (iStock)

NEWS CAREERS COMMENTARY JOURNALS Science brought to you by Vanderbilt Univer... LIBRARY LOG IN BECOME A MEMBER

News Home All News ScienceInsider News Features

2017

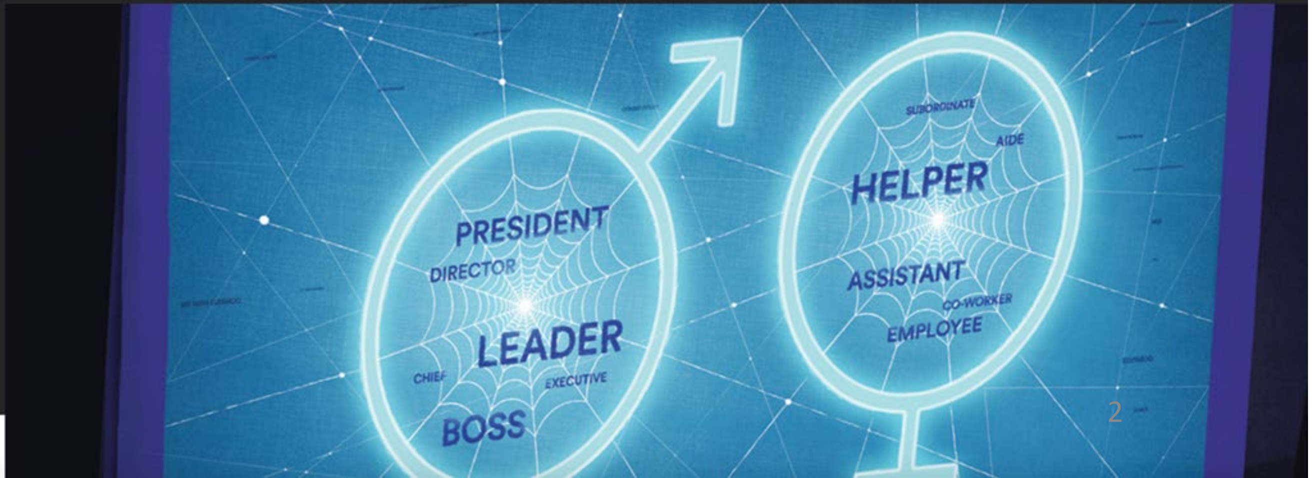
HOME > NEWS > ALL NEWS > EVEN ARTIFICIAL INTELLIGENCE CAN ACQUIRE BIASES AGAINST RACE AND GENDER

NEWS | TECHNOLOGY

Even artificial intelligence can acquire biases against race and gender

Computers can automatically adopt our biases by reading what we write

13 APR 2017 • BY MATTHEW HUTSON



A Lack of Variables Can Bias Machine Learning

Science | Current Issue | First release papers | Archive | About | Submit manusc

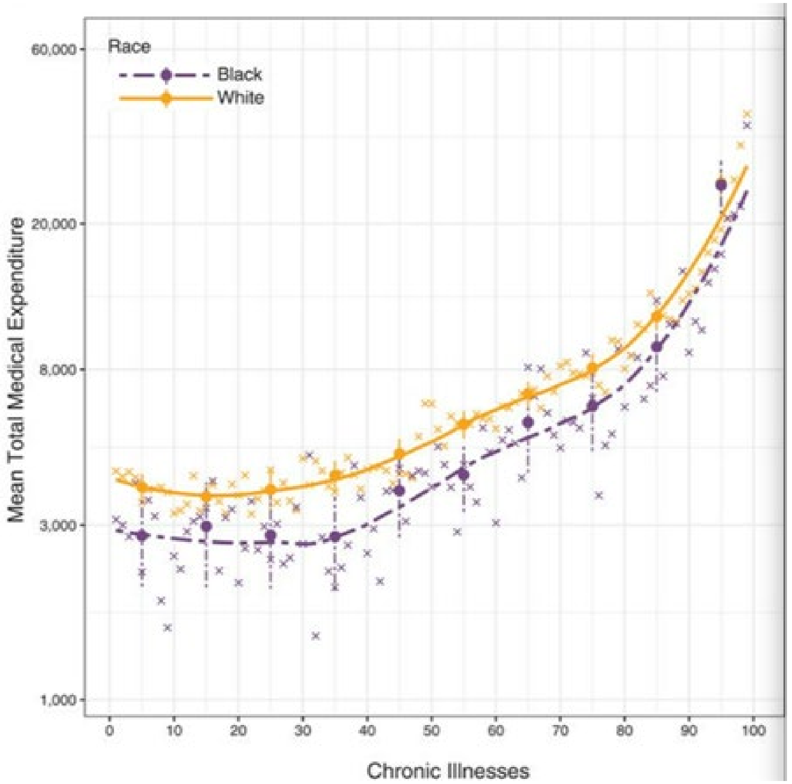
HOME > SCIENCE > VOL. 366, NO. 6464 > DISSECTING RACIAL BIAS IN AN ALGORITHM USED TO MANAGE THE HEALTH OF POPULATIONS

RESEARCH ARTICLE

Dissecting racial bias in an algorithm used to manage the health of populations

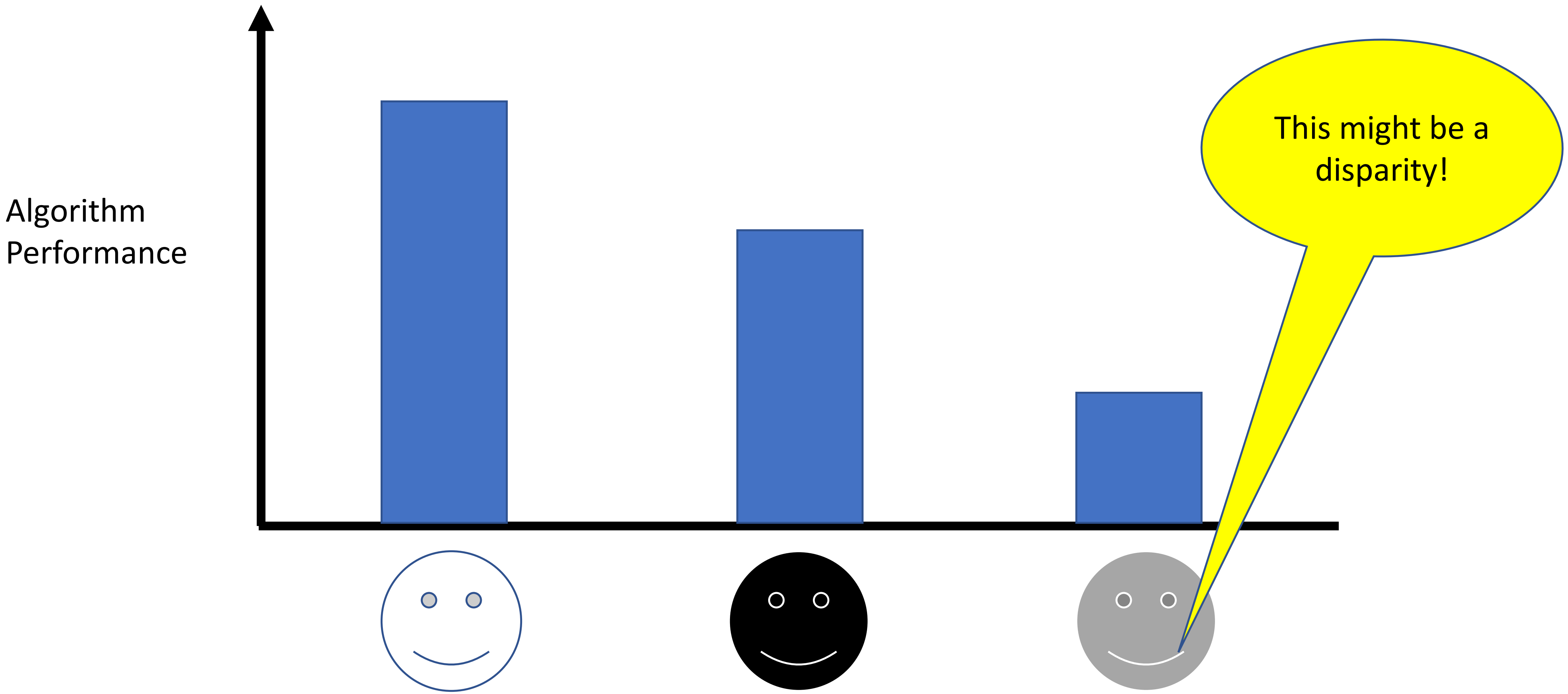
ZIAD OBERMEYER, BRIAN POWERS, CHRISTINE VOGELI, AND, SENDHIL MULLAINATHAN Authors Info & Affiliations

SCIENCE • 25 Oct 2019 • Vol 366, Issue 6464 • pp. 447-453 • DOI: 10.1126/science.aax2342

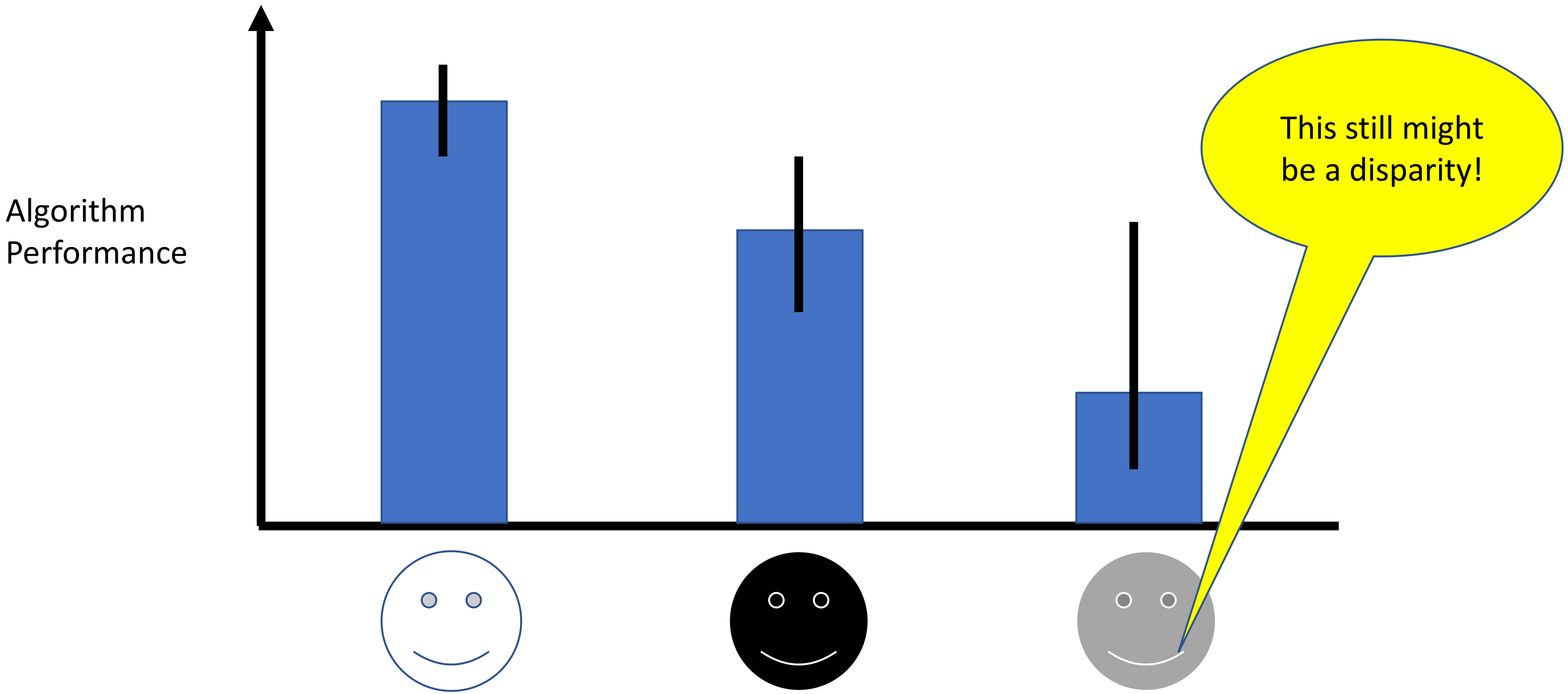


- The Problem
  - Payment levels were correlated with health outcomes
  - White patients paid more through insurance
  - The resulting machine learning model inferred that Whites needed more care than Blacks
- The Teachable Moment
  - One should check for correlations that obscure causality (the model should have included insurance status)

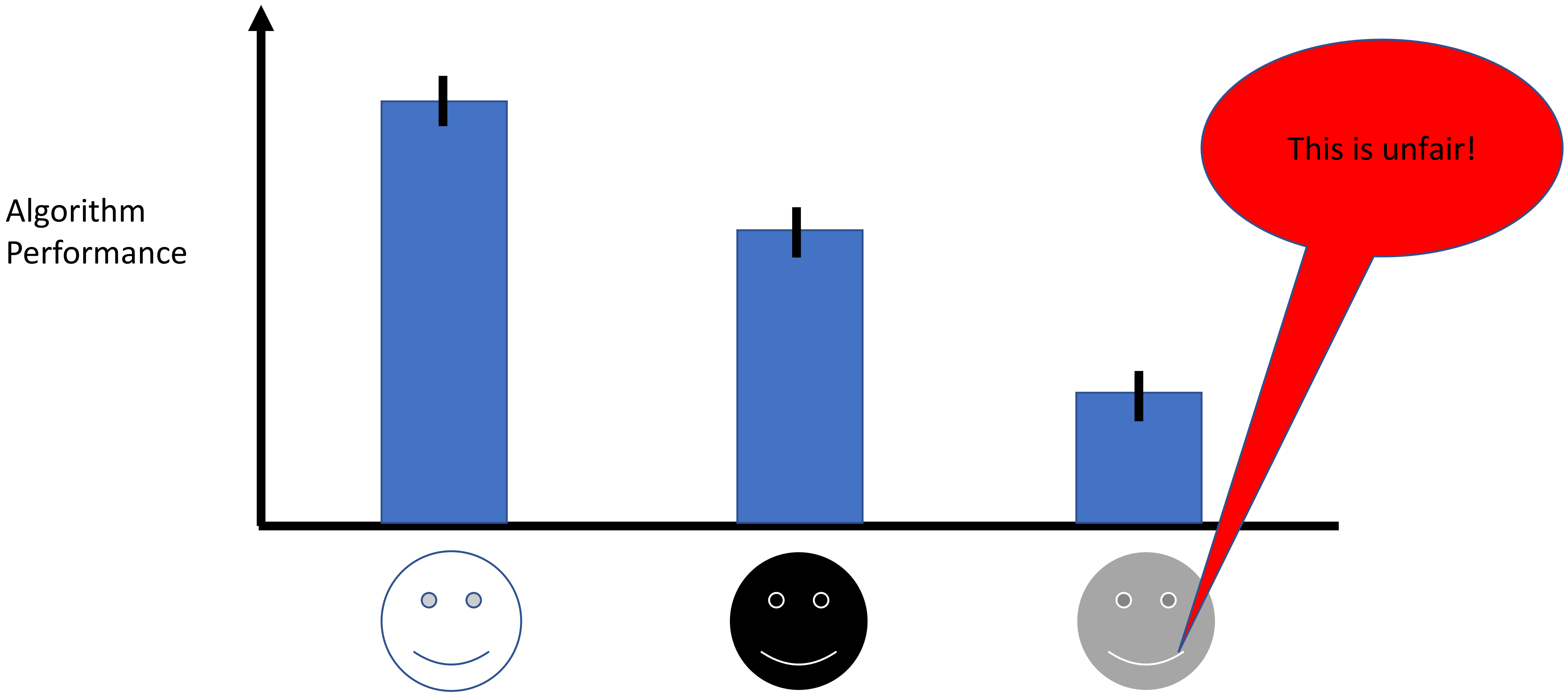
The “Fairness” Problem



The "Fairness" Problem



The “Fairness” Problem



Many Software Tools Test for Fairness

eBioMedicine
Part of THE LANCET *Discovery Science*
Volume 84, October 2022, 104250

Review

Algorithmic fairness in computational medicine

Jie Xu [a, b], Yunyu Xiao [b], Wendy Hui Wang [c], Yue Ning [c], Elizabeth A. Shenkman [a], Jiang Bian [a], Fei Wang [b]

Show more

+ Add to Mendeley   Share   Cite

https://doi.org/10.1016/j.ebiom.2022.104250   Get rights and content

Table 3. Popular library for fairness research.

Project Name	Developer	Description
FairMLHealth[81]	KenSci	Tools and tutorials for evaluating bias in healthcare machine learning.
AIF360[82]	IBM	Fairness metrics for datasets and machine learning algorithms, interpretation of the metrics, and approaches for reducing bias in datasets and models. It is available in both Python and R.
Fairlearn[83]	Microsoft	A Python package to evaluate fairness and mitigate any observed inequities. Fairlearn includes mitigation algorithms and metrics for model evaluation. It also contains Jupyter notebooks with examples of Fairlearn usage.
Fairness-comparison[84]	Sorelle et al.	Compare fairness-aware machine learning techniques. It aims to facilitate benchmarking of fairness-aware machine learning algorithms.
MEASURES[85]	Cardoso et al.	A benchmark framework for assessing discrimination-aware models.
Fairness Indicators[86]	Google	A suite of tools built on top of TensorFlow Model Analysis that enable regular evaluation of fairness metrics in product pipelines.
ML-fairness-gym[87]	Google	A general framework for studying and exploring long-term equity effects in carefully constructed simulation scenarios where learning subjects interact with the environment over time.
themis-ml[88]	Niels Bantilan	A Python library built on top of pandas and sklearn that implements fairness-aware machine learning algorithms.
FairML[89]	Julius Adebayo	A Python toolkit for auditing machine learning model deviations.

AI Fairness 360

Example

This extensible open source toolkit can help you examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle. We invite you to use and improve it.

Python API Docs ↗ | Get Python Code ↗ | Get R Code ↗

Not sure what to do first? Start here!

Read More

Learn more about fairness and bias mitigation concepts, terminology, and tools before you begin.

→

Try a Web Demo

Step through the process of checking and remediating bias in an interactive web demo that shows a sample of capabilities available in this toolkit.

→

Watch Videos

Watch videos to learn more about AI Fairness 360.

→

Read a paper

Read a paper describing how we designed AI Fairness 360.

→

Use Tutorials

Step through a set of in-depth examples that introduces developers to code that checks and mitigates bias in different industry and application domains.

→

Ask a Question

Join our AIF360 Slack Channel to ask questions, make comments and tell stories about how you use the toolkit.

→

View Notebooks

Open a directory of Jupyter Notebooks in GitHub that provide working examples of bias detection and mitigation in sample datasets. Then share your own notebooks!

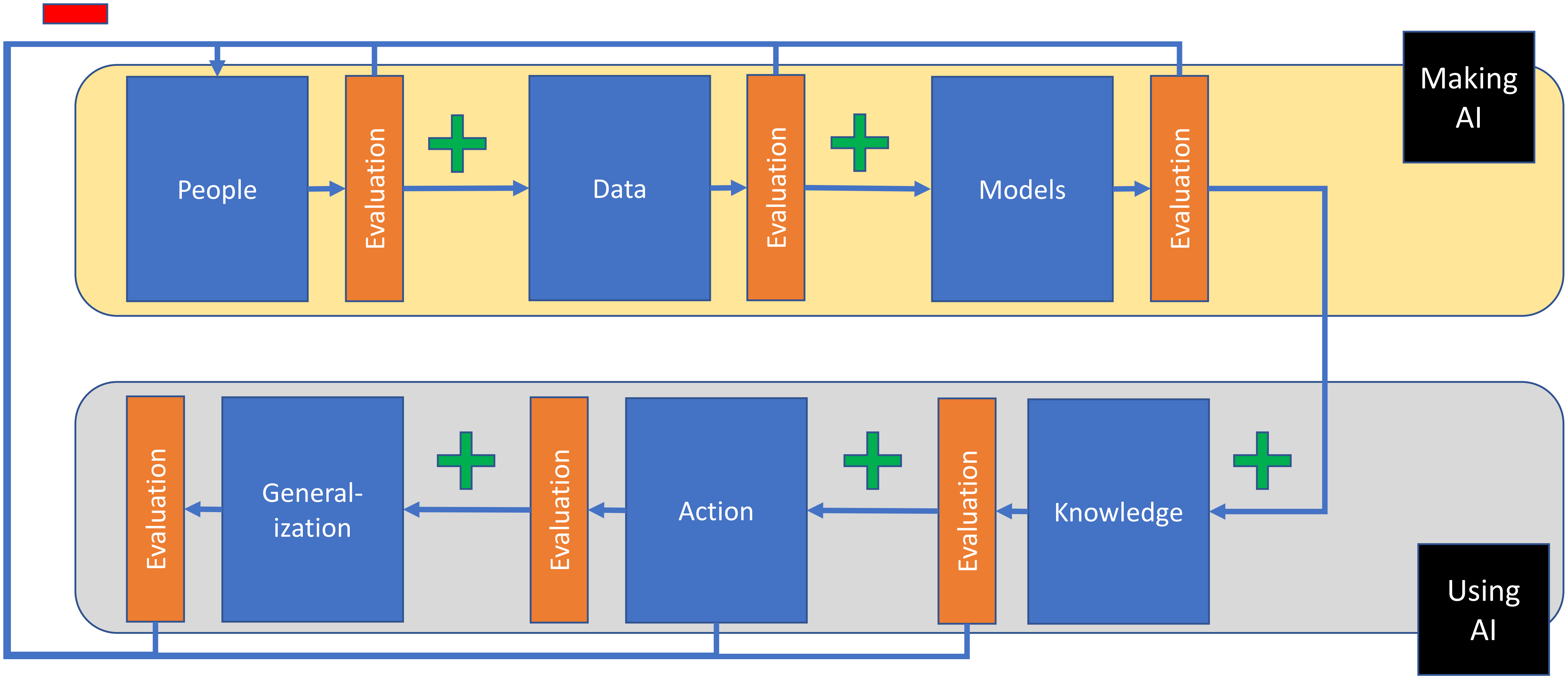
→

Contribute

You can add new metrics and algorithms in GitHub. Share Jupyter notebooks show-casing how you have examined and mitigated bias in your machine learning application.

→

Ethics Must be Embedded from the Outset

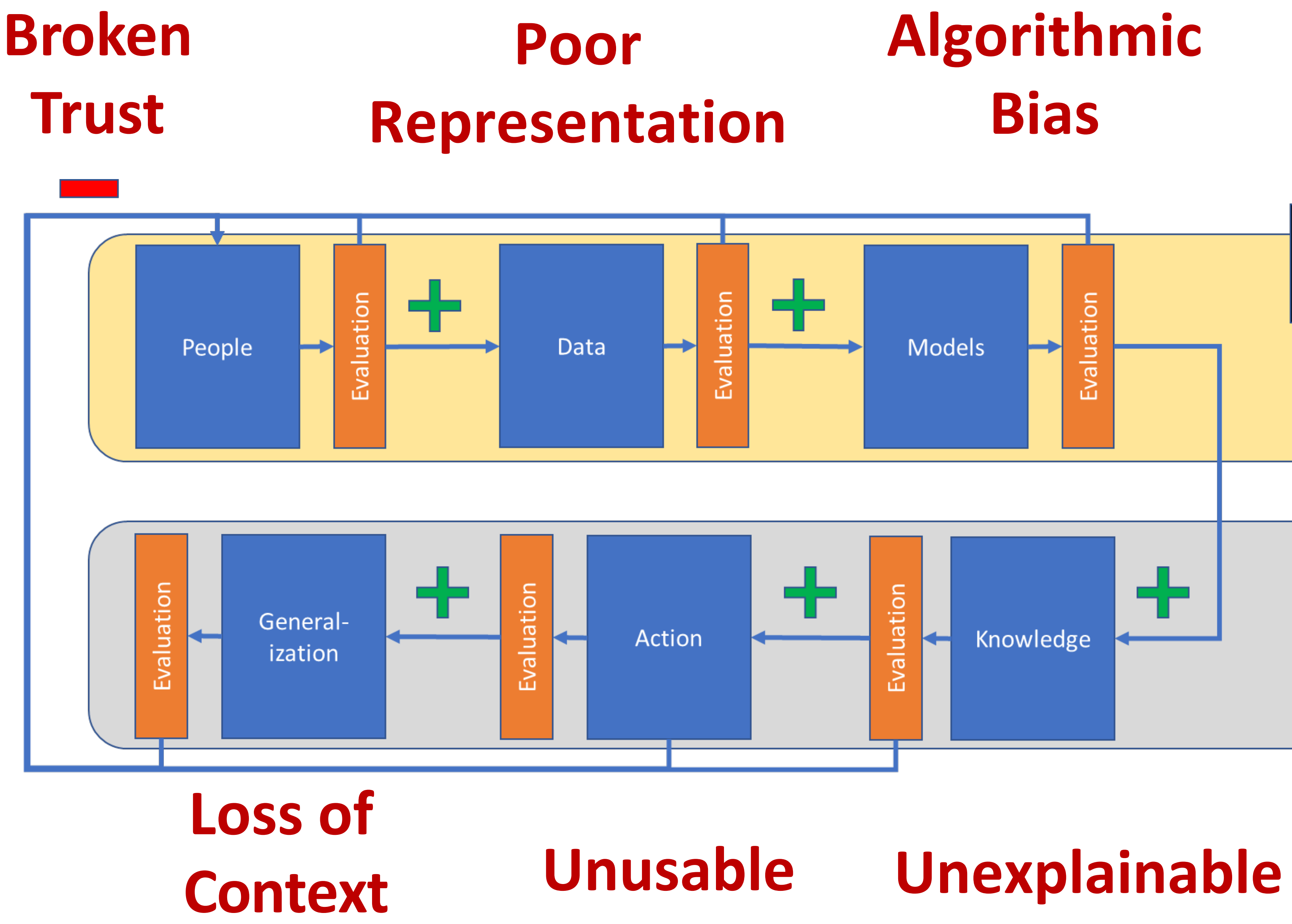


https://www.forbes.com/sites/glenngow/2021/07/11/google-facebook-and-microsoft-are-working-on-ai-ethics-heres-what-your-company-should-be-doing/

Ethics can help concretize the goals of AI/ML and to notify us to the pitfalls along the way

AI for Social Good

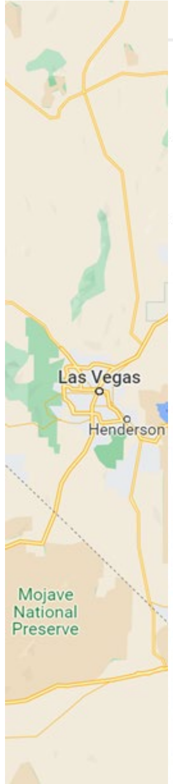
Image from AI for Social Good Workshop NeurIPS2018



Broken
Trust
Poor
Representation
Algorithmic
Bias
Making
AI
People
Evaluation
Data
Evaluation
Models
Evaluation
Evaluation
General-
ization
Evaluation
Action
Evaluation
Knowledge
Using
AI
Loss of
Context
Unusable
Unexplainable

Simply Because Data Exists, Doesn't Mean You Can Use it for Anything

- https://journalofethics.ama-assn.org/article/genetic-research-among-havasupai-cautionary-tale/2011-02
- The Problem
  - Data was collected by University of Arizona from the Havasupai tribe with consent for a certain purpose
  - The data was "reused" by university researchers for other purposes beyond the scope of consent
- The Teachable Moment
  - Data are about individuals, communities, and cultures. Data should not be used in a manner that disrespect expectations



The New York Times

Indian Tribe Wins Fight to Limit Research of Its DNA

321

Edmond Tilousi, 56, who can climb the eight miles to the rim of the Grand Canyon in three hours. Jim Wilson/The New York Times

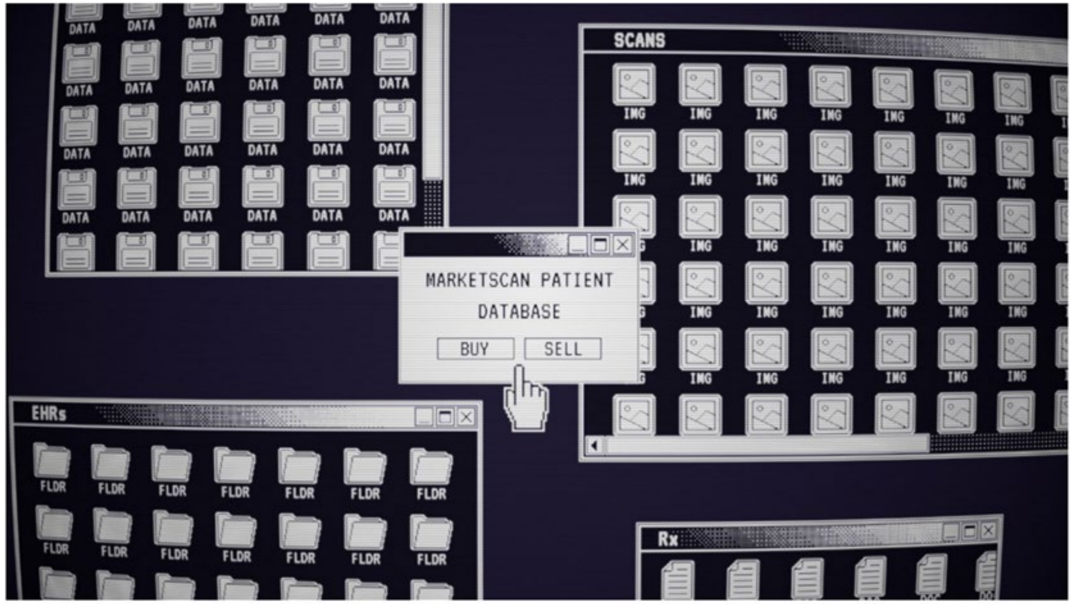
SPECIAL REPORT

STAT+

How a decades-old database became a hugely profitable dossier on the health of 270 million Americans

By Casey Ross Feb. 1, 2022

Reprints



ALEX HOGAN/STAT

THE WALL STREET JOURNAL.

Subscribe | Sign In

Home World U.S. Politics Economy Business Tech Markets Opinion Books & Arts Real Estate Life & Work WSJ. Magazine Sports

SHARE

◆ WSJ NEWS EXCLUSIVE | TECH

Google's 'Project Nightingale' Gathers Personal Health Data on Millions of Americans

Search giant is amassing health records from Ascension facilities in 21 states; patients not yet informed

Tech giants like Amazon and Apple are expanding their businesses to include electronic health records -- which contain data on diagnoses, prescriptions and other medical information. That's creating both opportunities and spurring privacy concerns. Here's what to know. Photo Composite: Heather Seidel/ The Wall Street Journal

By Rob Copeland

Updated Nov. 11, 2019 4:27 pm ET

PRINT TEXT 657

Google is engaged with one of the U.S.'s largest health-care systems on a project to collect and crunch the detailed personal-health information of millions of people across 21 states.

Only Use Models in Context

- https://jamanetwork.com/journals/jamainternalmedicine/article-abstract/2781313

- The Problem
  - The Epic EHR system developer trained a machine learning model to predict sepsis using a certain population's data
  - When the model was reused with a new population, the performance was substantially worse than the original results suggested
- The Teachable Moment
  - Models should not be used out of context. Know your populations!

STAT+ https://www.statnews.com/2022/10/24/epic-overhaul-of-a-flawed-algorithm/

SPECIAL REPORT

Epic's overhaul of a flawed algorithm shows why AI oversight is a life-or-death issue

By Casey Ross Oct. 24, 2022

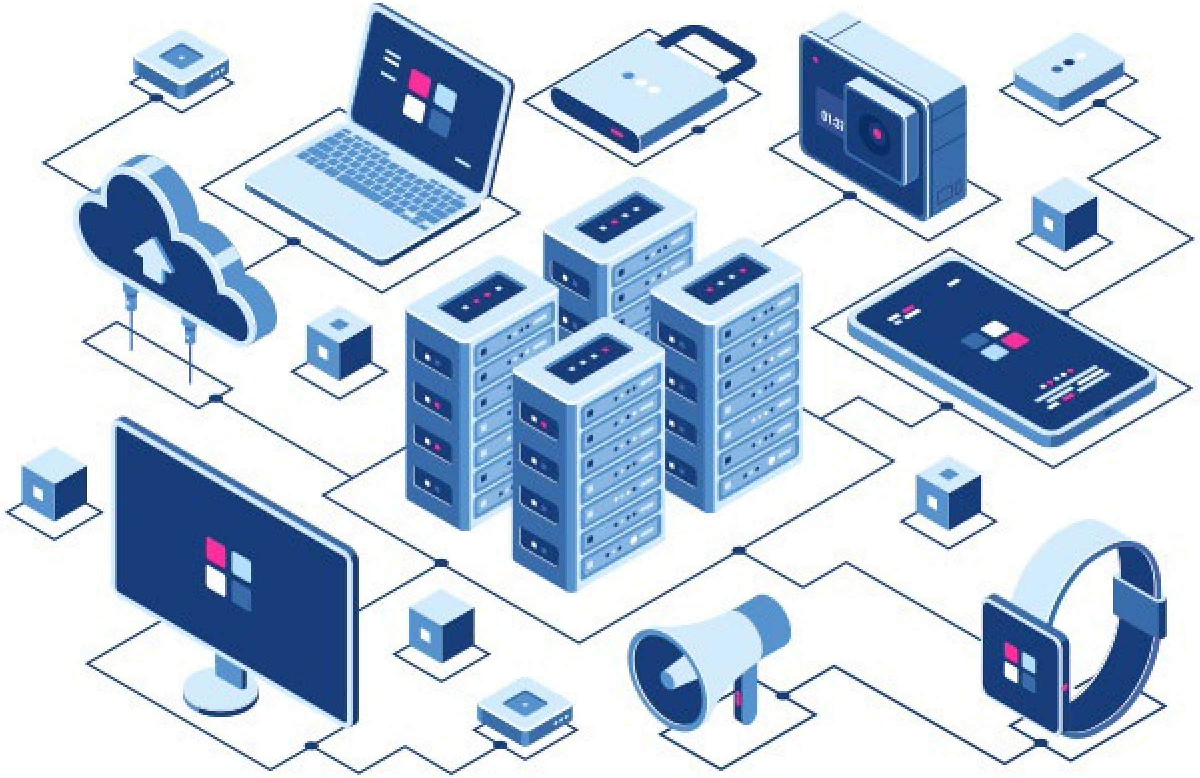
MOLLY FERGUSON FOR STAT

Reprints

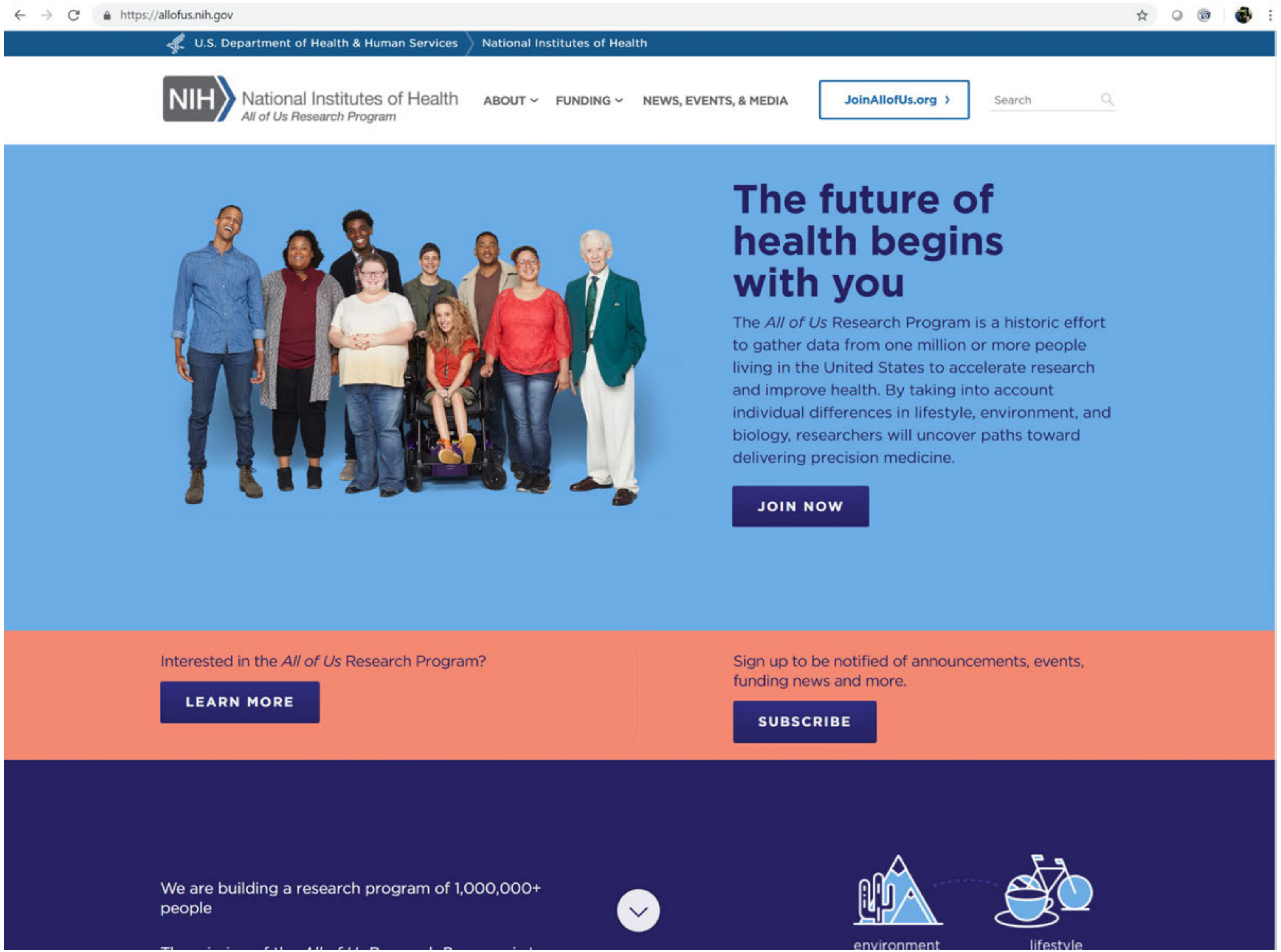
Epic, the nation's dominant seller of electronic health records, was bracing for a catastrophe.

It was June 2021, and a study about to be published in the Journal of the American Medical Association had found that Epic's artificial intelligence tool to predict sepsis, a deadly complication of infection, was prone to missing cases and flooding clinicians with false alarms. Reporters were clamoring for an

Infrastructure is Needed to Support Everything



Research Ready Environments Can Be Created



https://databrowser.researchallofus.org/

Search Across Data Types

Keyword Search

Data includes 331,360 participants and is current as of 11/29/2021.

FAQs

Introductory Videos

User Guide

EHR Domains

Conditions
23,300
medical concepts
201,920 participants in this domain
View Conditions

Drug Exposures
28,798
medical concepts
194,420 participants in this domain
View Drug Exposures

Labs & Measurements
14,502
medical concepts
199,040 participants in this domain
View Labs & Measurements

Procedures
27,444
medical concepts
185,580 participants in this domain
View Procedures

Genomics

Genomic Variants
98,000
participants in the Whole Genome Sequencing (WGS) dataset
164,180
participants in the Genotyping Array dataset
View Genomic Variants

Physical Measurements and Wearables

Physical Measurements
8
Physical Measurements
274,540 participants in this domain
Participants have the option to provide a standard set of physical measurements.
View Physical Measurements

Fitbit
4
Fitbit Measurements
11,700 participants in this domain
Fitbit data includes heart rate and activity summaries.
View Fitbit

Survey Questions

The Basics
28
questions available
331,360 participants in this domain
This survey includes participant demographic information.

Overall Health
21
questions available
331,360 participants in this domain
Survey includes information about how participants report levels of individual health.

Lifestyle
26
questions available
331,360 participants in this domain
Survey includes information on participant smoking, alcohol and recreational drug use.

Personal Medical History
465
questions available
114,460 participants in this domain
This survey includes information about past medical history, including medical conditions and approximate age of

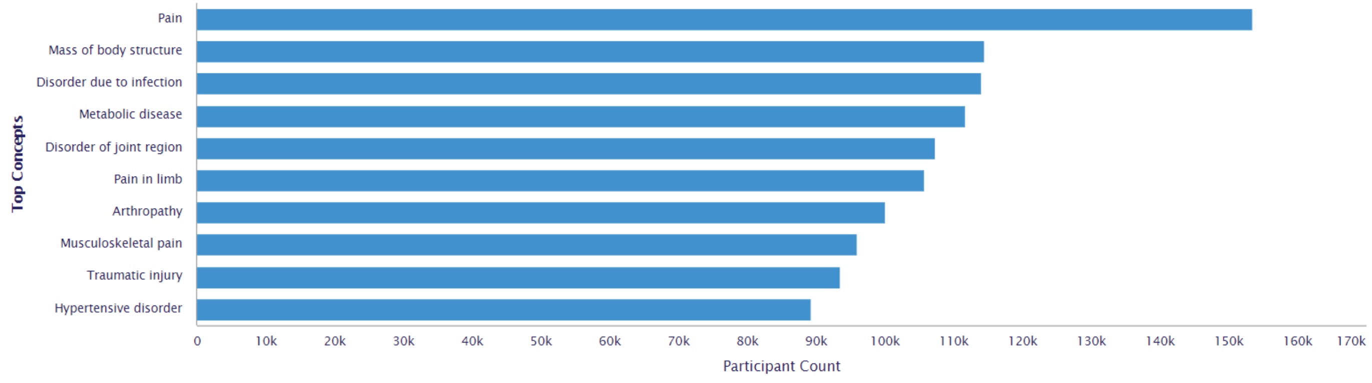
https://databrowser.researchallofus.org/ehr/conditions

Keyword Search

< BACK TO MAIN SEARCH

DATA DISCLAIMER

Top 10 by Descending Participant Counts



Showing top 1 - 50 of 24770 concepts for this domain

Conditions	Participants of 192,000	% of 192,000
1. Pain *Also Known As* Pain (finding), Part hurts, Painful, Dolor, Pain observations	153,440	79.92%
2. Mass of body structure *Also Known As* Mass, Lump, Observation of a mass, Mass of body structure (finding)	114,460	59.61%
3. Disorder due to infection *Also Known As* Infectious disease (disorder), Infective disorder, Disease due to infection, Infection, Infectious d... See More	113,900	59.32%
4. Metabolic disease *Also Known As* Metabolic disorder, MD - Metabolic disorders, Metabolic disease (disorder)	111,580	58.11%
5. Disorder of joint region	107,280	55.87%

https://www.researchallofus.org/data-tools/workbench/

All of Us Research Hub | NIH National Institutes of Health *All of Us Research Program*

ABOUT DATA & TOOLS DISCOVER FAQ REGISTER

Researcher Workbench

Researcher Workbench

The Researcher Workbench is a cloud-based platform where registered researchers can access Registered and Controlled Tier data. Its powerful tools support data analysis and collaboration. Integrated help and educational resources are provided through the Workbench User Support Hub.

WORKSPACES

Registered researchers use workspaces to access, store, and analyze data for specific research projects. Workspaces are collaborative and can be shared among other registered researchers within a project team.

USES: Organizing research projects, collaboration

Workspaces Preview >

NOTEBOOKS

Researchers with R or Python experience can perform high-powered queries and analysis within the *All of Us* datasets using our integrated, cloud-based Jupyter Notebook environment.

USES: Analysis, queries

Notebooks Preview >

DATASET BUILDER

The Dataset Builder allows researchers to search and save

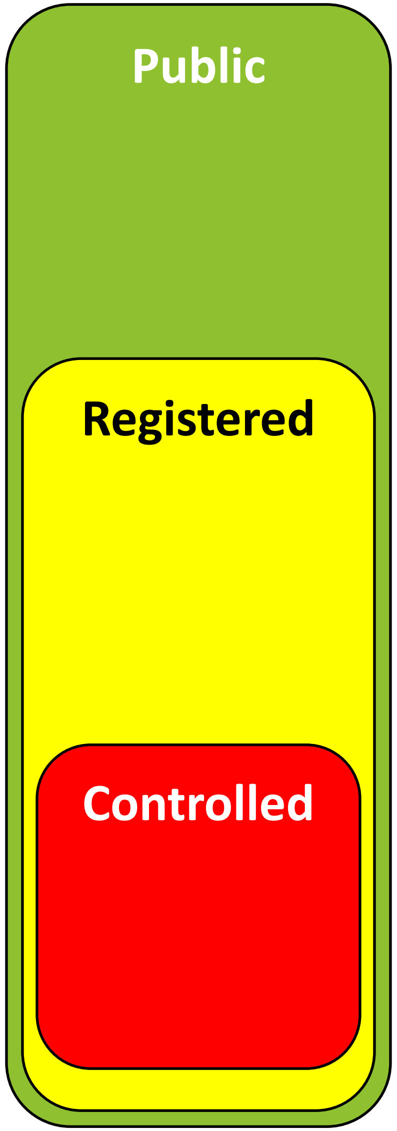
COHORT BUILDER

The Cohort Builder is a custom, point-and-click tool that allows

Tiered Levels of Access in All of Us Research Program

- Public
  - Can be accessed without logging in
  - Summary statistics only

- Sandbox Environments
  - Registered
    - Available to anyone within a trusted organization… plans to expand out to citizen scientists
    - Individual-level data with low risk of re-identification
  - Controlled – released earlier this year (with 100k human genomes)!
    - Available to researchers in a trusted organizations
    - More detail, more risk, but still designated as non-human subjects



Engaging a Diverse Researcher Community

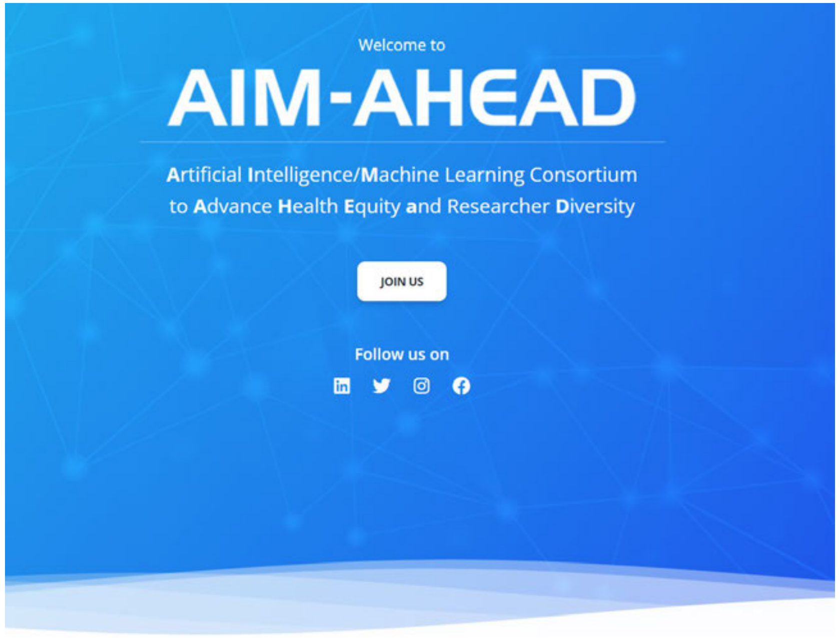
Dr. Watson, Dr. Kitani Lemieux, and Jaelyn Stepter at Xavier University of Louisiana.

Jaelyn was the first place winner of this year's Minority Student Research Symposium.

How *All of Us* Engages with Diverse Researcher Communities:

- **Creating a pipeline for students**: The *All of Us* Minority Student Research Symposium (MSRS)
- **Partnerships with HBCUs through CPGI Network**: Xavier University of Louisiana
- **Partnership with Baylor College of Medicine**: *All of Us* Evenings with Genetics Research Program series with Dr. Debra Dianne Murray

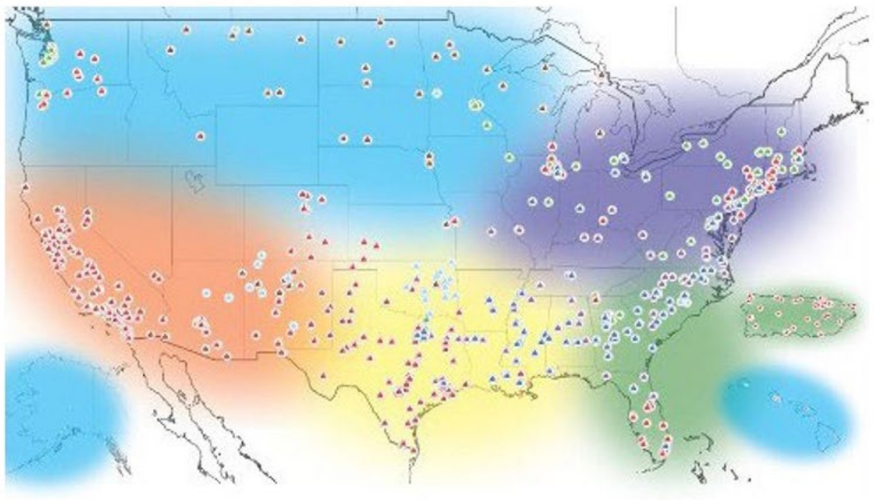
https://aim-ahead.net/



AIM AHEAD

The National Institutes of Health's Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity (AIM-AHEAD) program has established mutually beneficial, coordinated, and trusted partnerships to enhance the participation and representation of researchers and communities currently underrepresented in the development of AI/ML models and to improve the capabilities of this emerging technology, beginning with electronic health records (EHR) and extending to other diverse data to address health disparities and inequities.

The AIM-AHEAD Program, a Hub and Spoke Model



https://aim-ahead.net/

Leadership Core
University of North Texas Health Science Center in Fort Worth

Regional Hubs
Vanderbilt University Medical Center
University of Houston
University of North Texas Health Science Center in Fort Worth
University of Colorado-Anschutz Medical Center in Aurora
University of California, Los Angeles
Meharry Medical College in Nashville, Tennessee
Morehouse School of Medicine in Atlanta, Georgia
Johns Hopkins University in Baltimore, Maryland

Data Science Training Core
- Howard University in Washington, D.C.

Infrastructure Core
- National Alliance Against Disparities in Patient Health in Woodbridge, Virginia
- Harvard Medical School in Boston, Massachusetts
- Vanderbilt University Medical Center in Nashville, Tennessee

Data and Research Core
- OCHIN in Portland, Oregon

- ~20 research pilots beginning at institutions around the country at various HBCUs and MSIs
    - Addressing ethics and equity at various stages in the lifecycle
- ~20 research fellows
- ~20 leadership fellows

Learning About How We Learn

- Embedding ethnographers with researchers and research teams to
  - Gain intuition into how they collect data
  - Work with IRBs and administrative leadership to get projects up and running
  - Access data and infrastructure to perform machine learning
  - Determine needs and gaps for conducting meaningful AI development and implementation

The R’s at the Heart of Data for AI

- All about teams and experimental environments

- **Repeat**ability: Same Team, Same Experimental Setup
  - You can achieve the same result with the same data

- **Replica**bility: Different Team, Same Experimental Setup
  - Someone else can achieve the same result with the same data

- **Reproducibility**: Different Team, Different Experimental Setup
  - Someone else can achieve the same result with different data
  - Generalizable knowledge

Replicability – Data Sharing

- Journals have pushed for data sharing

PLOS ONE

PUBLISH ABOUT BROWSE SEARCH

advanced search

Introduction

Minimal Data Set Definition

Acceptable Data Sharing Methods

Acceptable Data Access Restrictions

Unacceptable Data Access Restrictions

FAQs

PLOS Data Advisory Board

Data Availability

The following policy applies to all PLOS journals, unless otherwise noted.

Introduction

PLOS journals require authors to make all data necessary to replicate their study's findings publicly available without restriction at the time of publication. When specific legal or ethical restrictions prohibit public sharing of a data set, authors must indicate how others may obtain access to the data.

When submitting a manuscript, authors must provide a Data Availability Statement describing compliance with PLOS' data policy. If the article is accepted for publication, the Data Availability Statement will be published as part of the article.

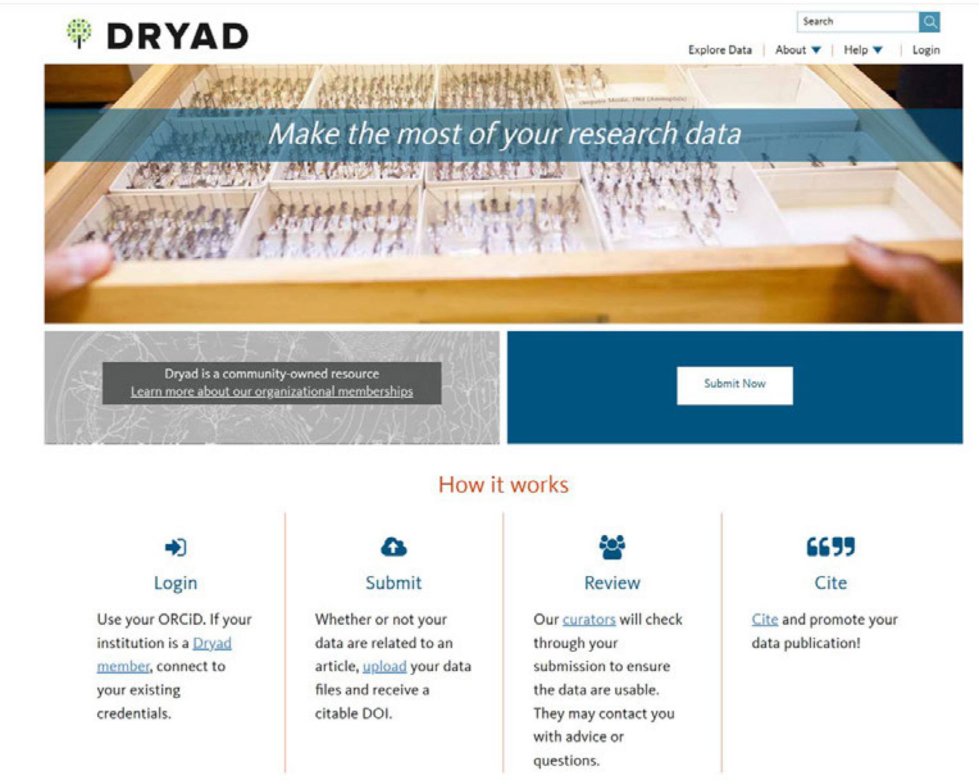
Acceptable data sharing methods are listed below, accompanied by guidance for authors as to what must be included in their Data Availability Statement and how to follow best practices in research reporting.

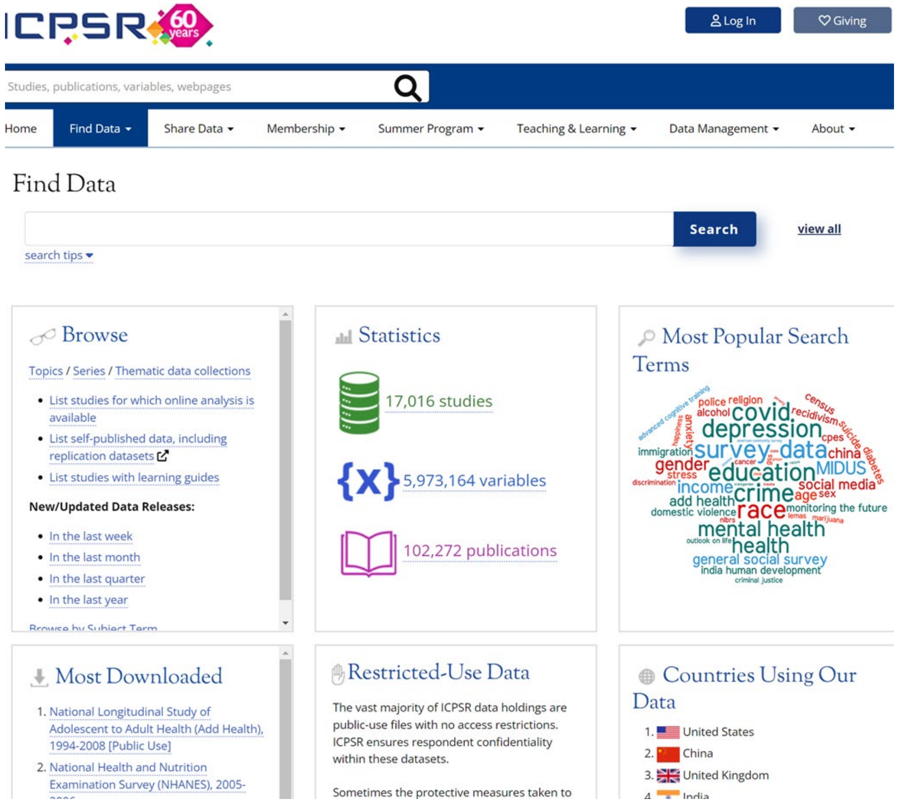
PLOS believes that sharing data fosters scientific progress. Data availability allows and facilitates:

- Validation, replication, reanalysis, new analysis, reinterpretation or inclusion into meta-analyses;

Replicability – Data Sharing

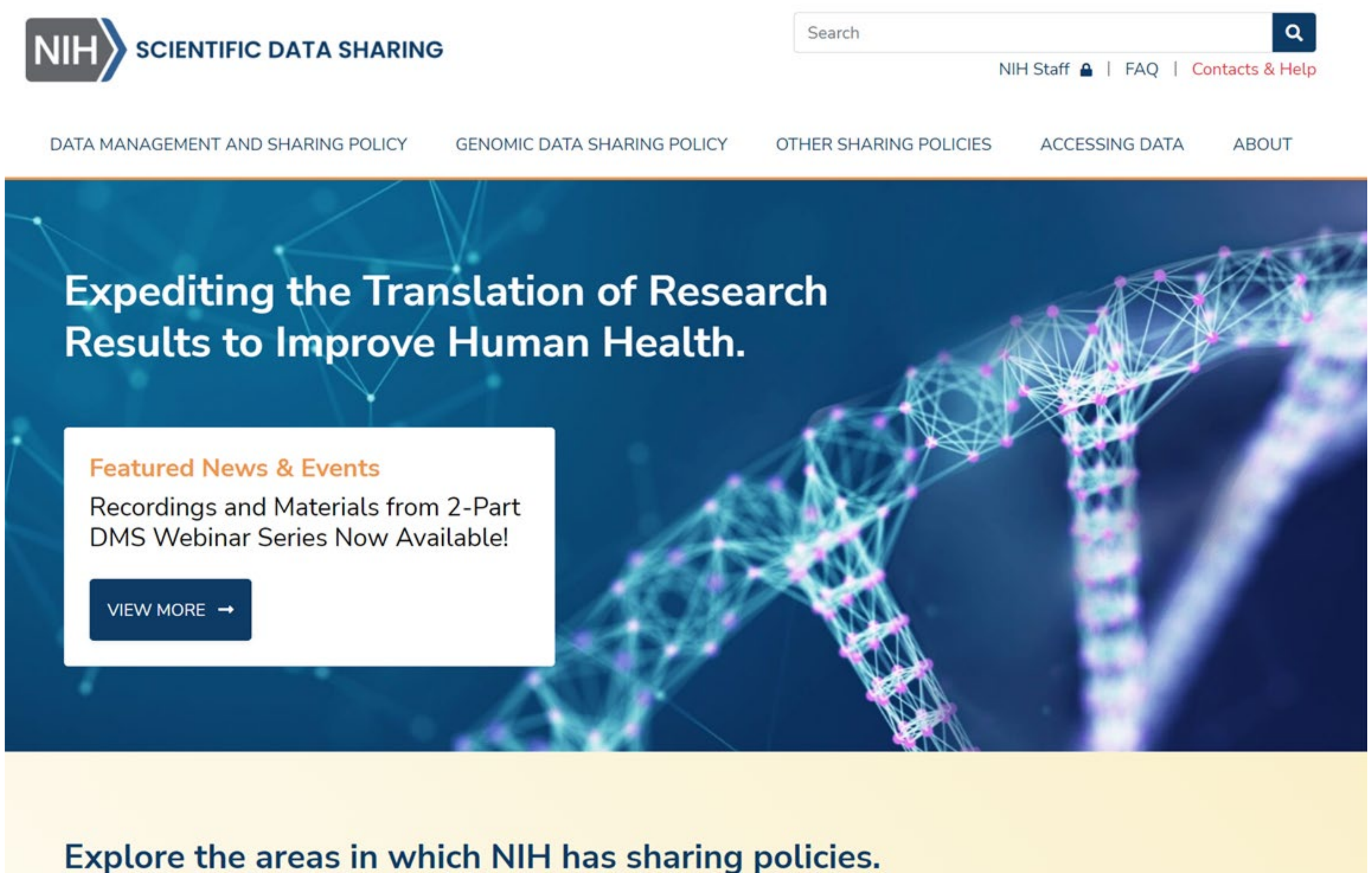
- Various repositories for data sharing have been established





Replicability – Data Sharing

- Various policies for data sharing have been established as well



Replicability – Data Sharing

- Journals have pushed for data sharing… **but**

PLOS ONE

PUBLISH ABOUT BROWSE SEARCH

advanced search

- Introduction
- Minimal Data Set Definition
- Acceptable Data Sharing Methods
- Acceptable Data Access Restrictions
- Unacceptable Data Access Restrictions
- FAQs
- PLOS Data Advisory Board

Data Availability

The following policy applies to all PLOS journals, unless otherwise noted.

Introduction

PLOS journals require authors to make all data necessary to replicate their study's findings publicly available without restriction at the time of publication. When specific legal or ethical restrictions prohibit public sharing of a data set, authors must indicate how others may obtain access to the data.

When submitting a manuscript, authors must provide a Data Availability Statement describing compliance with PLOS' data policy. If the article is accepted for publication, the Data Availability Statement will be published as part of the article.

Acceptable data sharing methods are listed below, accompanied by guidance for authors as to what must be included in their Data Availability Statement and how to follow best practices in research reporting.

PLOS believes that sharing data fosters scientific progress. Data availability allows and facilitates:

- Validation, replication, reanalysis, new analysis, reinterpretation or inclusion into meta-analyses;

Replicability – Data Sharing Pushback

- Numerous arguments, but most common invoked is privacy
- A problem that persists in human subjects research and data derived from the clinical domain (e.g., clinical trials or electronic health records)
- Numerous approaches to de-identification have been developed, but ensuring they are applied in practice has been a challenge

De-identification is Potentially Problematic

- Typically applied to hide (or amend) features that can be leveraged to identify an individual

- But the smaller the subpopulation, the more likely that a record will have information (e.g., geographic area, race, sexual orientation) amended in some way

- This can have major implications on bias and replicability

Self-Disclosure is a Big Problem

- De-identification often assumes that patients do not disclose their participation… but this is definitely not the case*

- And, disclosure can be made by the research program as well!

- This means that research programs must ask what their obligations are when offering privacy problem**

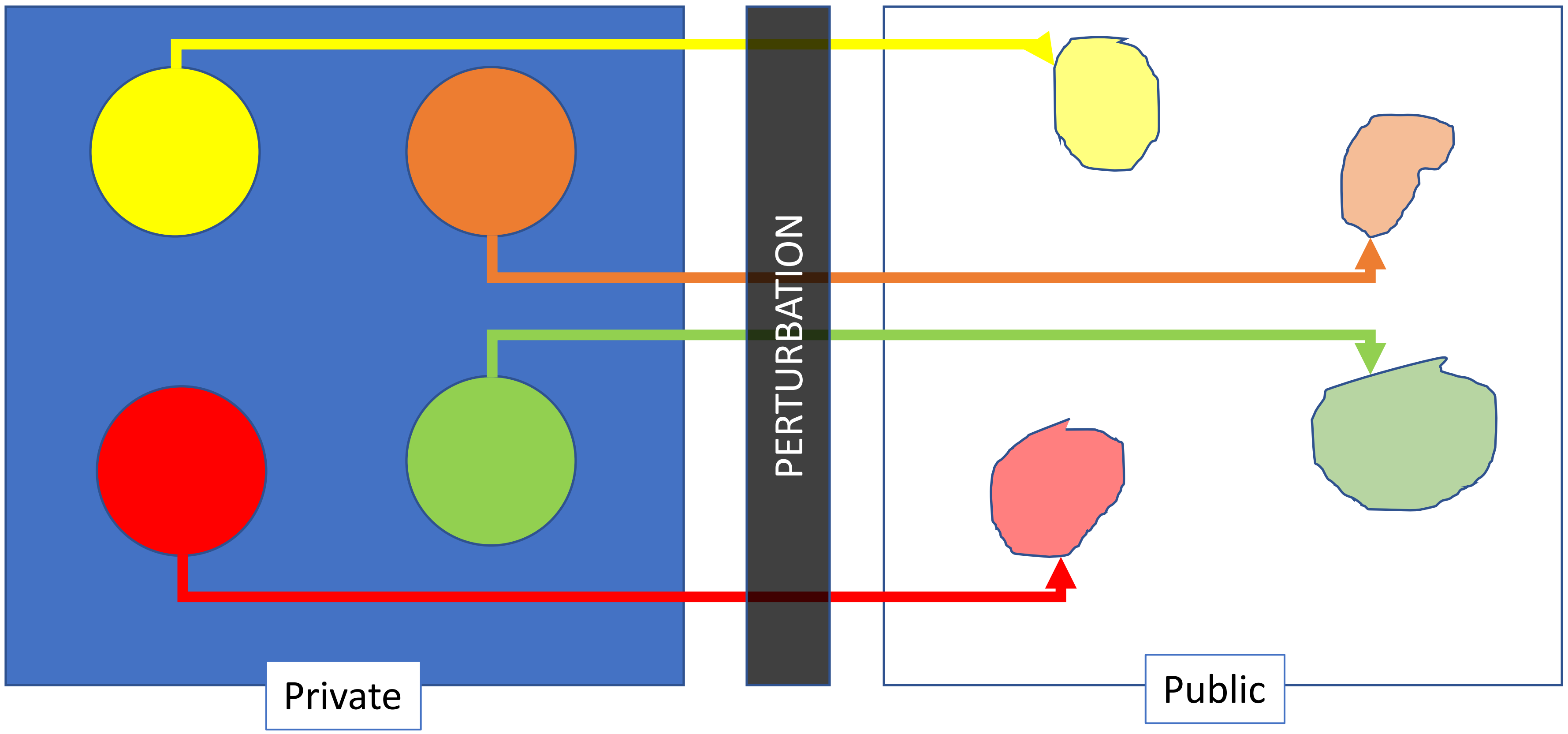
*Liu, et al. Biomedical research cohort membership disclosure on social media. AMIA. 2019.
**McKibbin, Malin, Clayton. Protecting research data of publicly revealing participants. Journal of Law and Biosciences. 2021.

Synthetic Data to the Rescue?

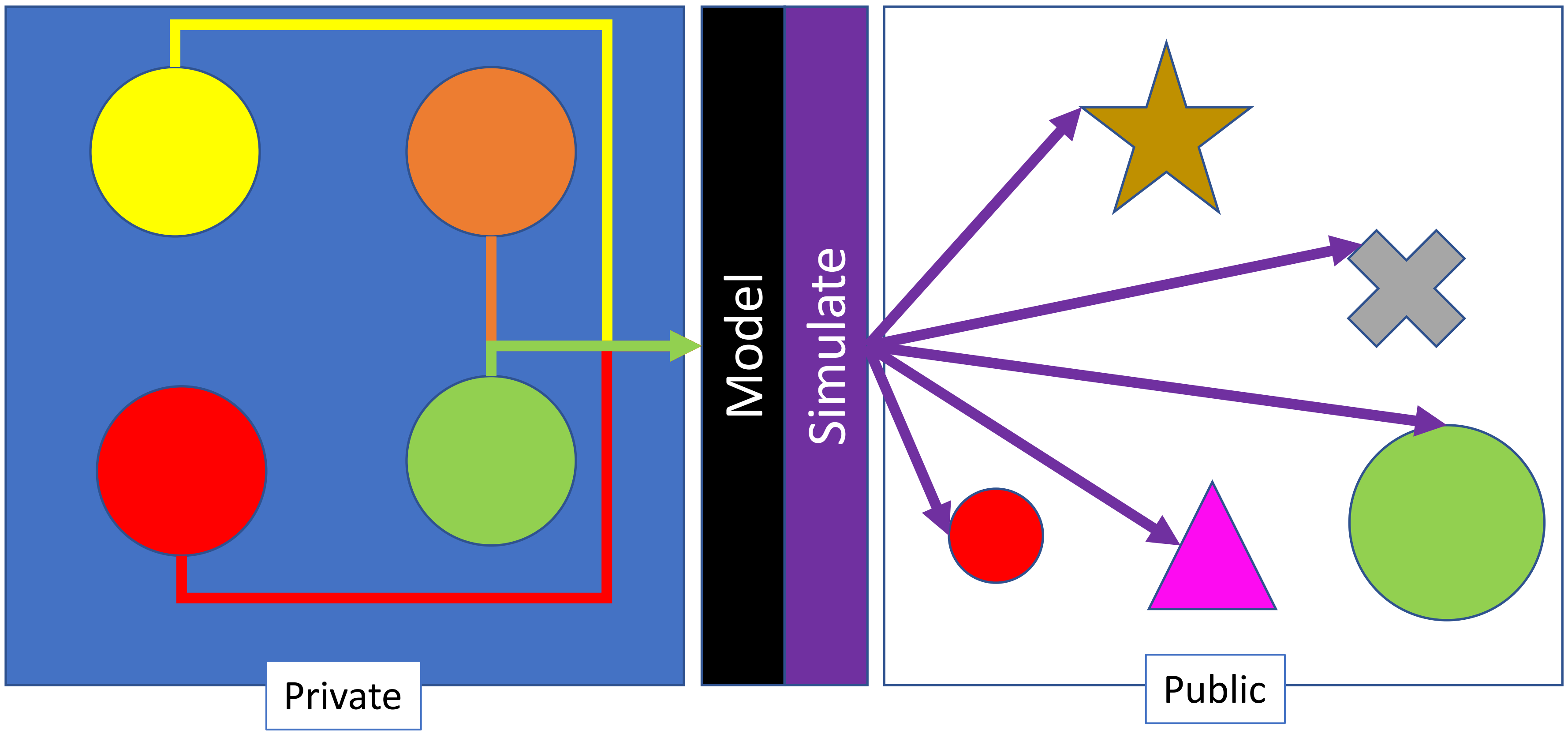
- Algorithmic bias often happens when there’s insufficient data on one population

- Can we “make” records for them?

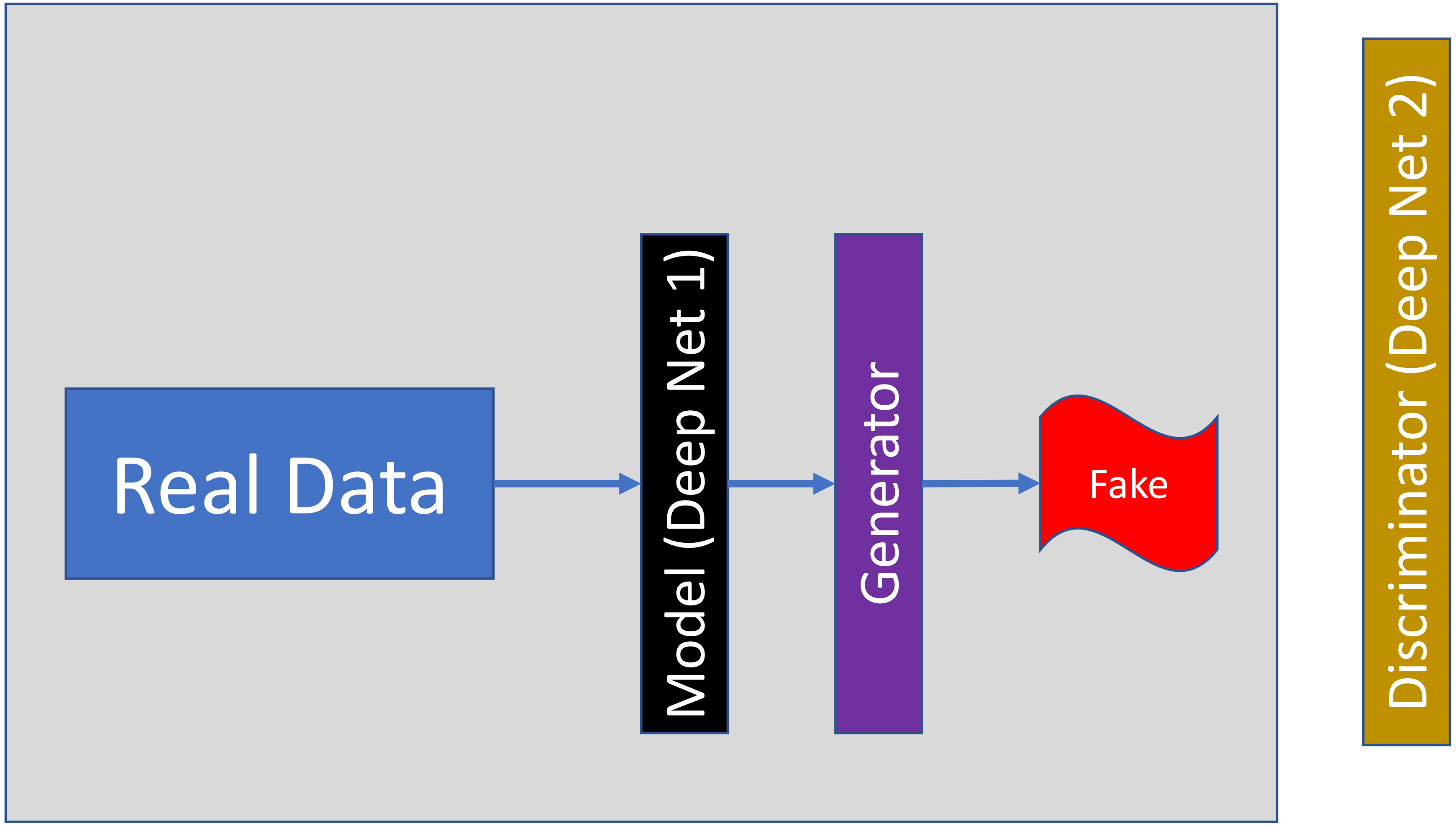
Ways to Generate Synthetic Data: Perturbation



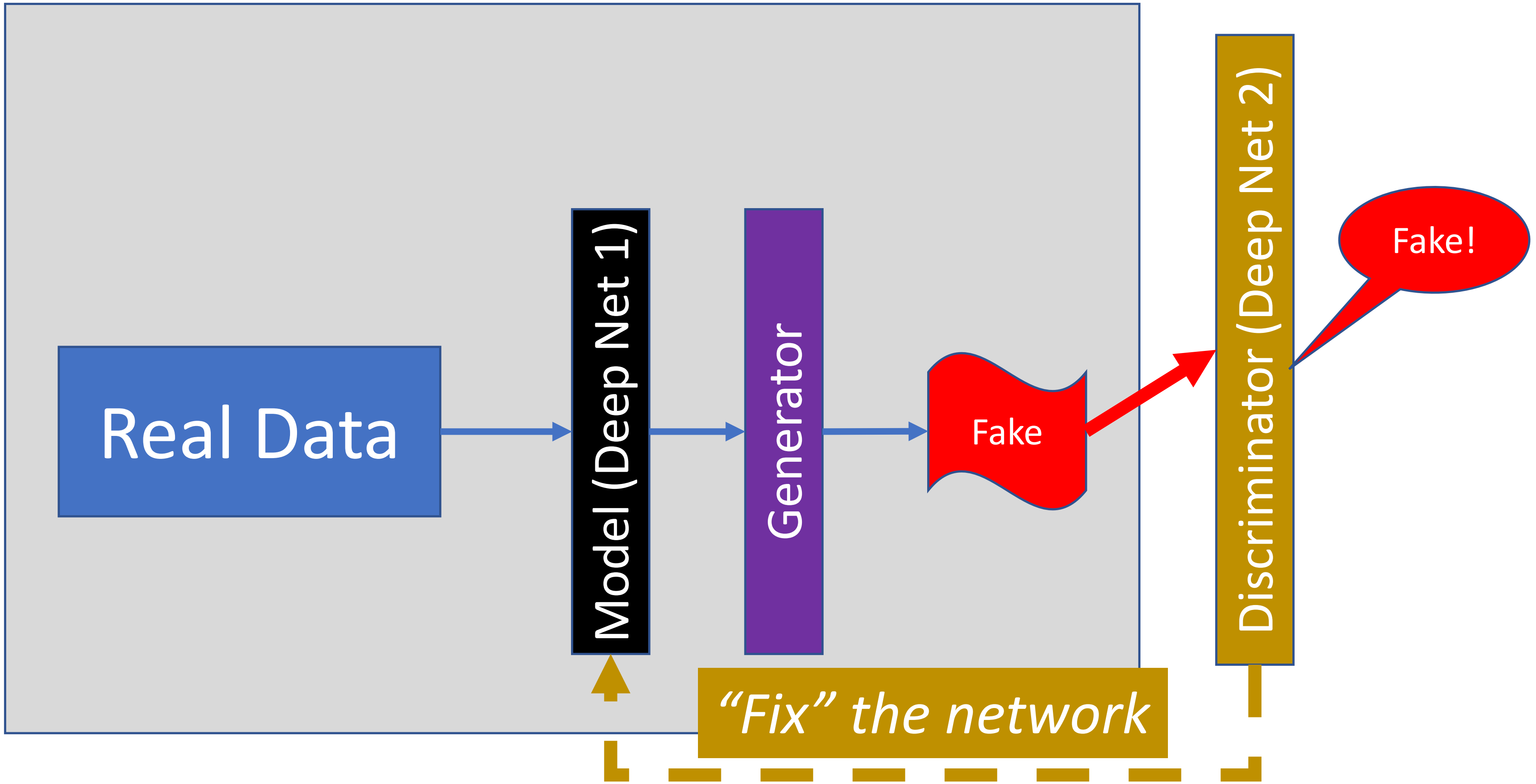
Ways to Generate Synthetic Data: Simulation



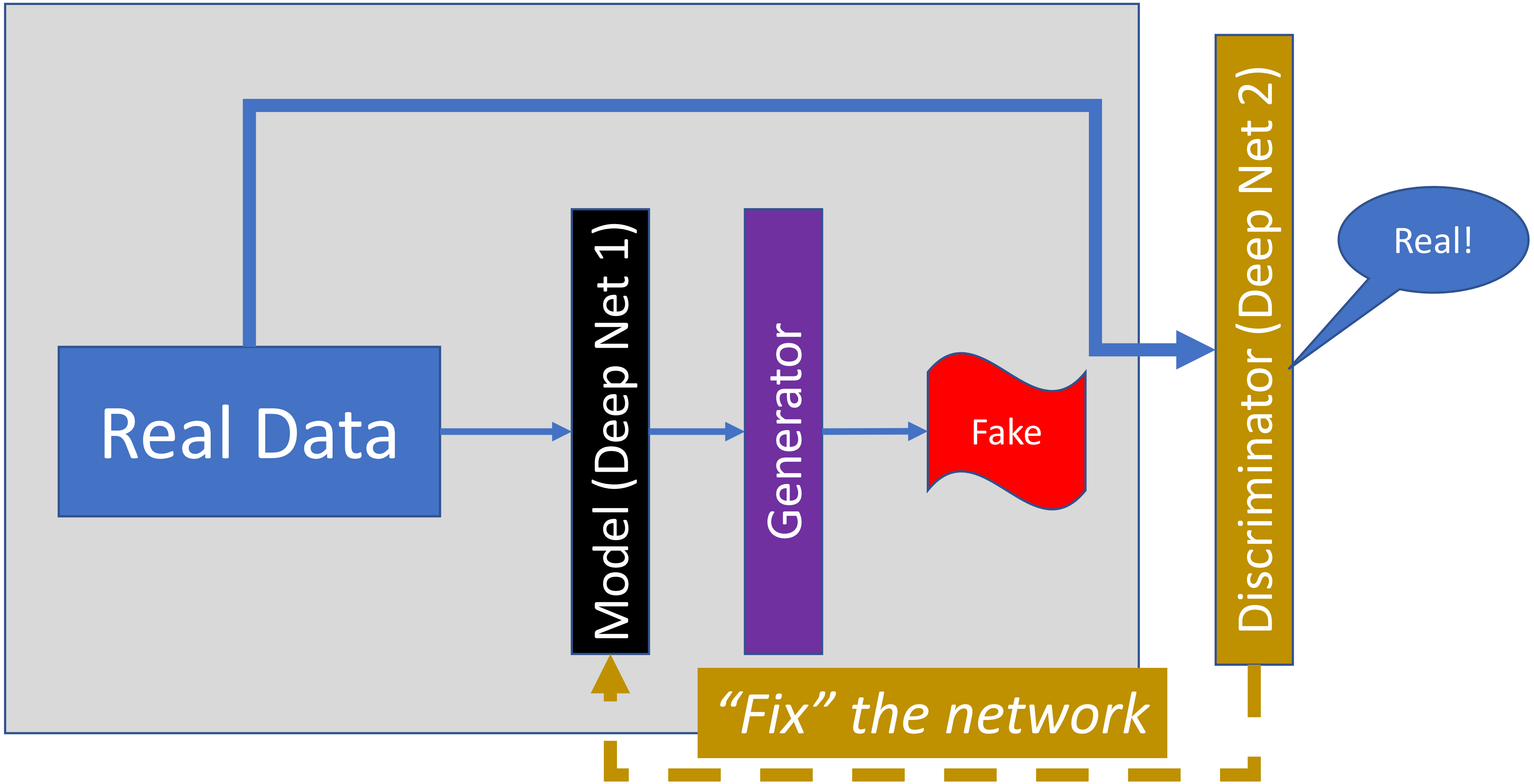
Generative Adversarial Networks: GANs



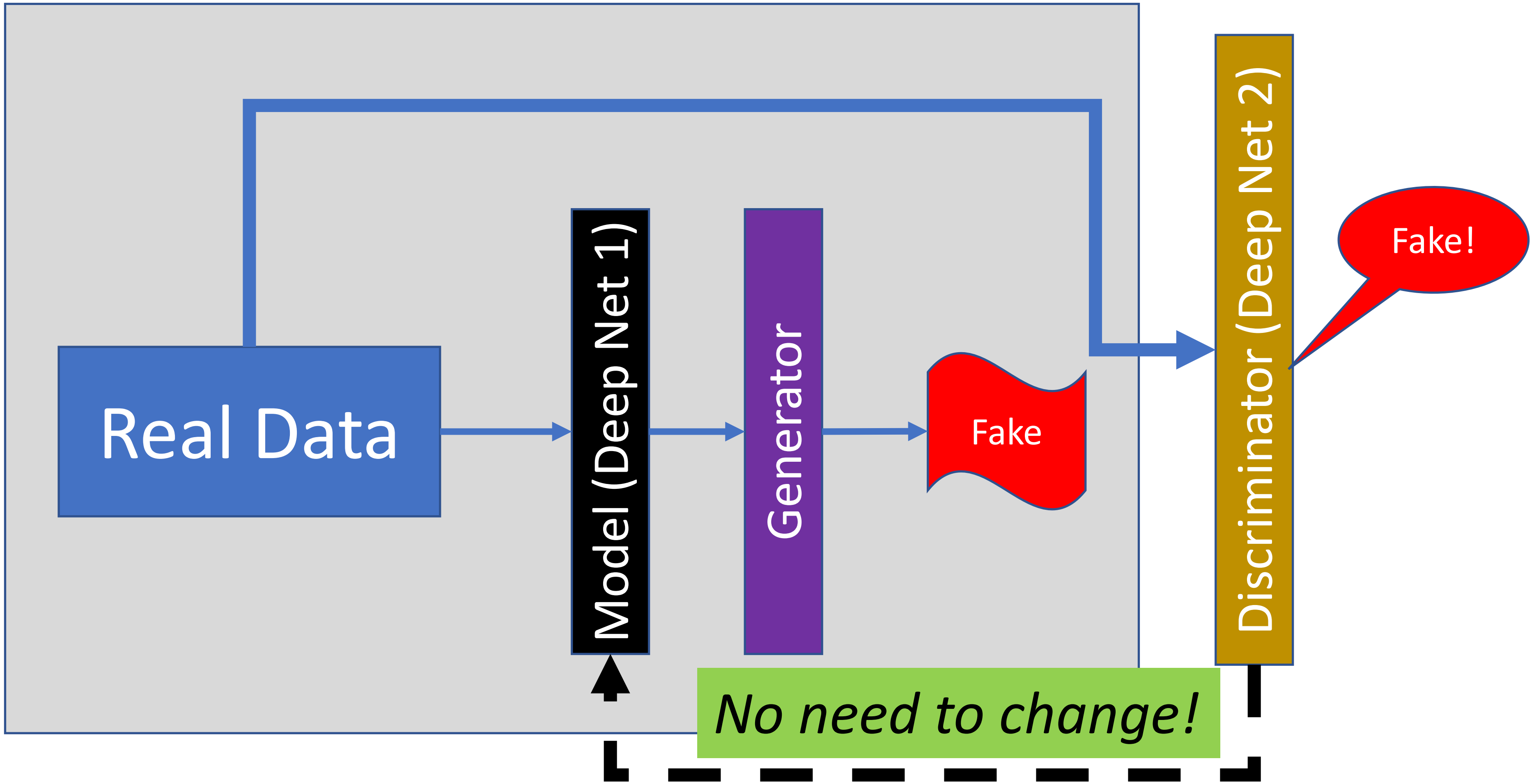
Generative Adversarial Networks: GANs



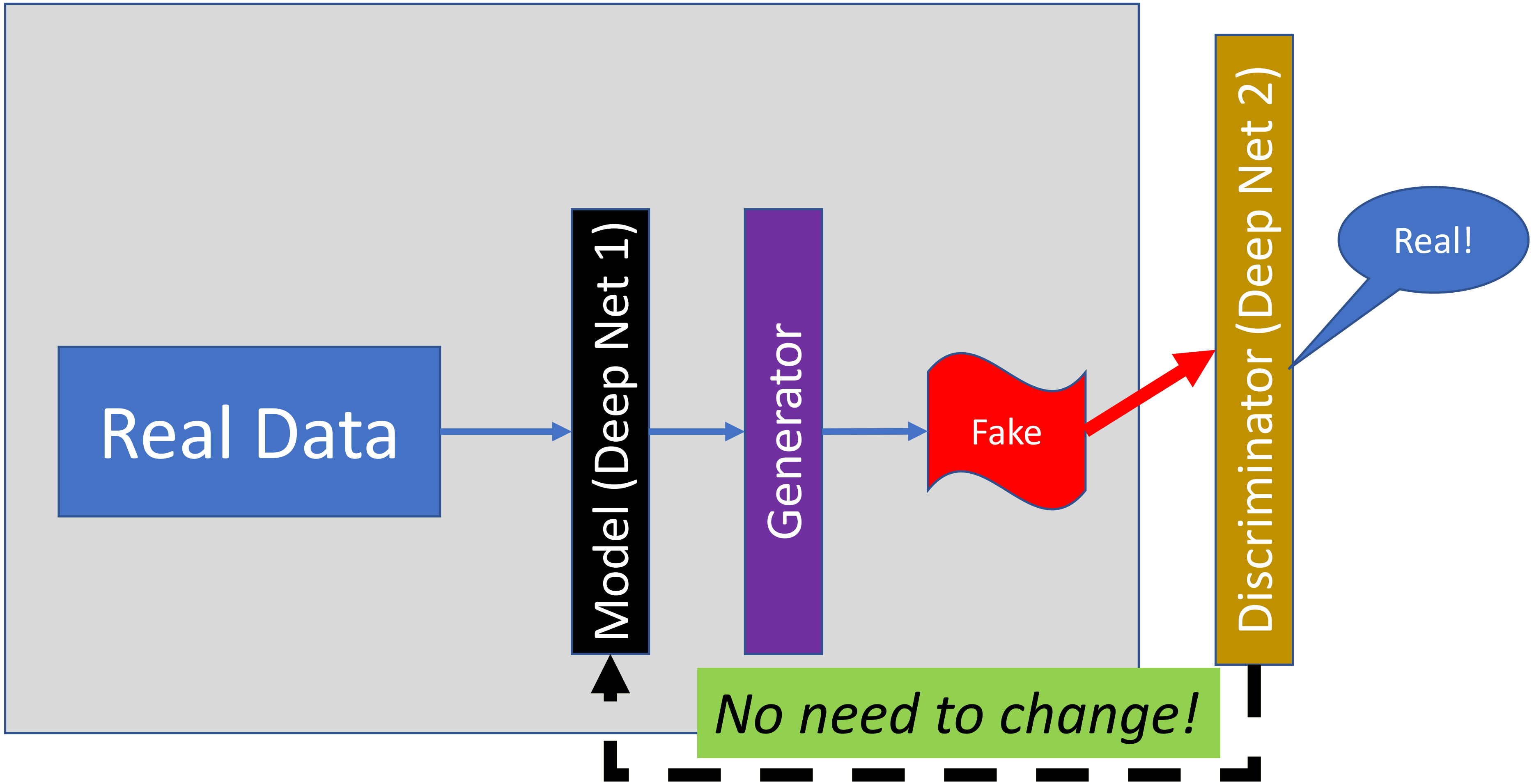
Generative Adversarial Networks: GANs



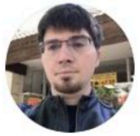
Playing the GAN Game



Playing the GAN Game



This is Not a New Principle



Ian Goodfellow
@goodfellow_ian

4.5 years of GAN progress on face generation.
arxiv.org/abs/1406.2661 arxiv.org/abs/1511.06434
arxiv.org/abs/1606.07536 arxiv.org/abs/1710.10196
arxiv.org/abs/1812.04948

Journal of Official Statistics, Vol. 18, No. 4, 2002, pp. 531–543

Satisfying Disclosure Restrictions With Synthetic Data Sets

Jerome P.

To avoid disclosures, Rubin proposed creating r
so that (i) no unit in the released data has sensiti
and (ii) statistical procedures that are valid for th
In this article, I show through simulation stu
from synthetic data in a variety of settings, inc
proportional to size sampling, two-stage clust
provide guidance on specifying the number and
the benefit of including design variables in the

Key words: Confidentiality; disclosure; multipl

JPC Privacy Confidentiality

Current Archives Announcements TPDP workshop Submissions

Home / Archives / Vol. 1 No. 1 (2009): Inaugural Issue / **Articles**

Estimating Risks of Identification Disclosure in Partially Synthetic Data

PDF

Published: Apr 1, 2009

DOI: https://doi.org/10.29012/jpc.v1i1.567

Keywords: Confidentiality Public use data

Jerome P. Reiter
Department of Statistical Science, Duke Uni
https://orcid.org/0000-0002-8374-383

Robin Mitra
University of Southampton, Southampton, U
https://orcid.org/0000-0001-9584-804

Abstract

To limit disclosures, statistical agencies and

Journal of Official Statistics, Vol. 28, No. 4, 2012, pp. 583–590

Inferentially Valid, Partially Synthetic Data: Generating from Posterior Predictive Distributions not Necessary

Jerome P. Reiter[1] *and Satkartar K. Kinney*[2]

To avoid disclosures in public use microdata, one approach is to release partially synthetic data sets. These comprise the units originally surveyed with some collected values, for example sensitive values at high risk of disclosure or values of key identifiers, replaced with multiple imputations. In practice, partially synthetic data typically are generated from Bayesian posterior predictive distributions; that is, one draws repeated values of parameters in the synthesis models before generating data from them. We show, however, that inferentially valid, partially synthetic data can be generated by fixing the parameters of the synthesis models at their modes. We do so with both a theoretical example and illustrative simulation studies. We also discuss implications of these results for agencies generating synthetic data.

Key words: Confidentiality; disclosure; imputation; microdata; privacy; survey.

This is Not a New Principle

(Choi et al MLHC 2017)

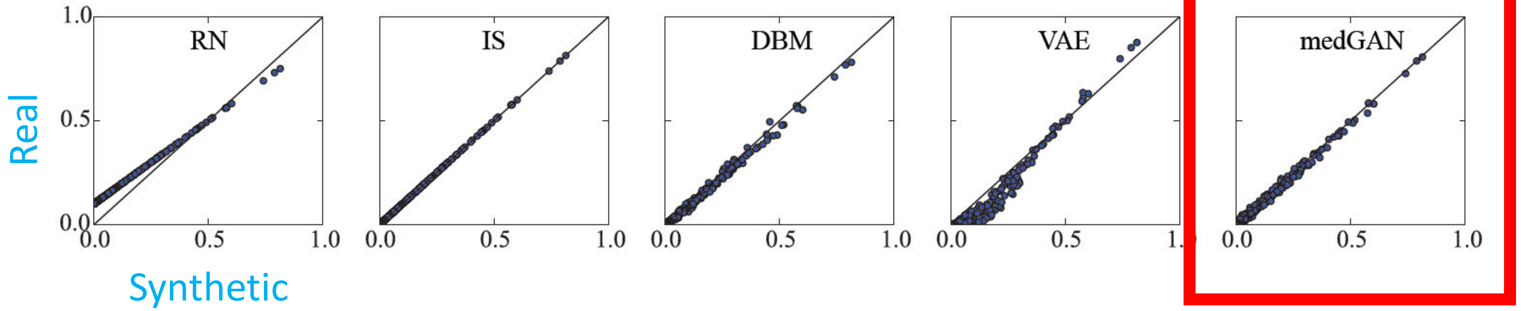
Proceedings of Machine Learning for Healthcare 2017 JMLR W&C Track Volume 68

Generating Multi-label Discrete Patient Records using Generative Adversarial Networks

Edward Choi[1] MP2893@GATECH.EDU
Siddharth Biswal[1] SBISWAL7@GATECH.EDU
Bradley Malin[2] BRADLEY.MALIN@VANDERBILT.EDU
Jon Duke[1] JON.DUKE@GATECH.EDU
Walter F. Stewart[3] STEWARWF@SUTTERHEALTH.ORG
Jimeng Sun[1] JSUN@CC.GATECH.EDU

[1]Georgia Institute of Technology [2] Vanderbilt University [3] Sutter Health

- Sutter Health & MIMIC
- Demographics, Diagnoses, Procedures, & Meds
- Prediction of presence / absence clinical concept



Evolution

- Better training (Wasserstein distance) and evaluation methods (latent dimensions) (Zhang et al JAMIA 2020)
- Enabling constraints (e.g., preventing women from having prostate cancer) (Yan et al AMIA 2020)
- Move from static to longitudinal data: think LSTMs + GANs (Zhang et al JAMIA 2021)

Zhang, Yan, Mesa, Sun, & Malin. Ensuring electronic medical record simulation through better training, modeling, and evaluation. JAMIA. 2020; 27: 99-108.
Yan, Zhang, Nyemba, & Malin. Generating electronic health records with multiple data types and constraints. Proc AMIA Symp. 2020: 1335-1344.
Zhang, Yan, Lasko, Sun, & Malin. SynTEG: A framework for temporal structured electronic health data simulation. JAMIA. 2021; 28: 596-604.

Case Study for Demos & Tutorial

> 30 researcher outreach and training events

> 2000 users

Researcher Workbench launched

1 year after Launch

May 2020

May 2021

All of Us
RESEARCH PROGRAM

Building a Synthetic Resource

Data selection

235,000 participants w/
demographics, physical measures, survey responses

Model training

Train multiple **GAN** models
to determine the best converge point and control repetition

Data generation

Based on **specific requirements** (e.g., volume, constraints)

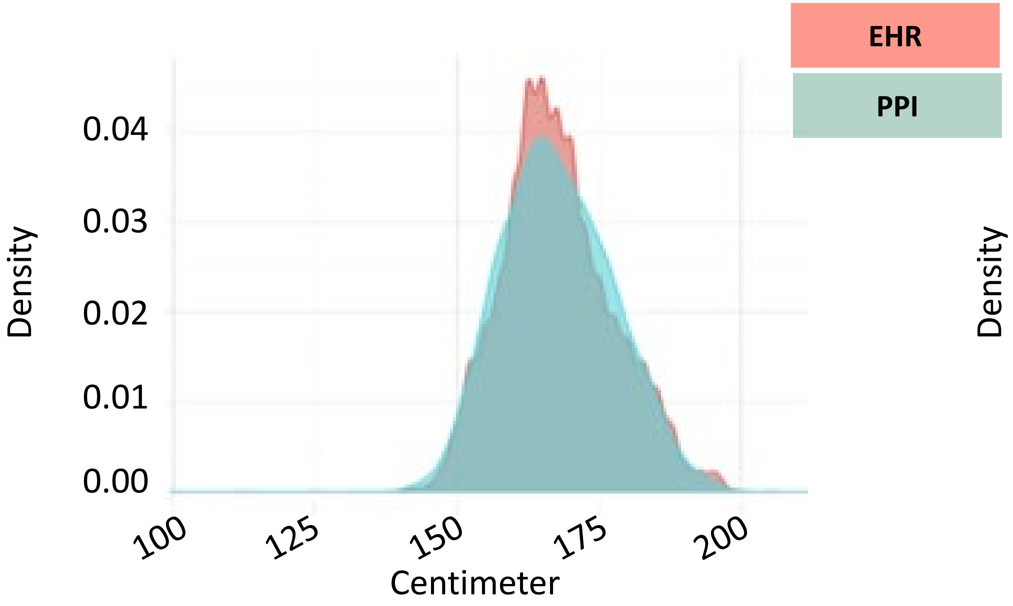
Post-hoc curation

Reinsert important outliers that are discarded by the data generation model, e.g. Wolfram syndrome

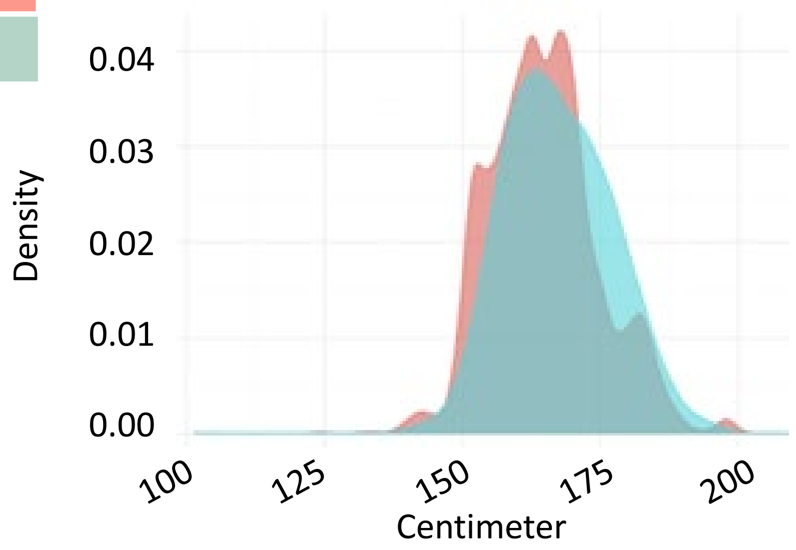
Two datasets to be made public later this year

All of Us
RESEARCH PROGRAM

Real *vs* Synthetic in the Same Tutorial



Using ***real*** data in RW



Using ***synthetic*** data in mirror RW

What Could Go Wrong?

ARTIFICIAL INTELLIGENCE

AI fake-face generators can be rewound to reveal the real faces they trained on

Researchers are calling into doubt the popular idea that deep-learning models are "black boxes" that reveal nothing about what goes on inside

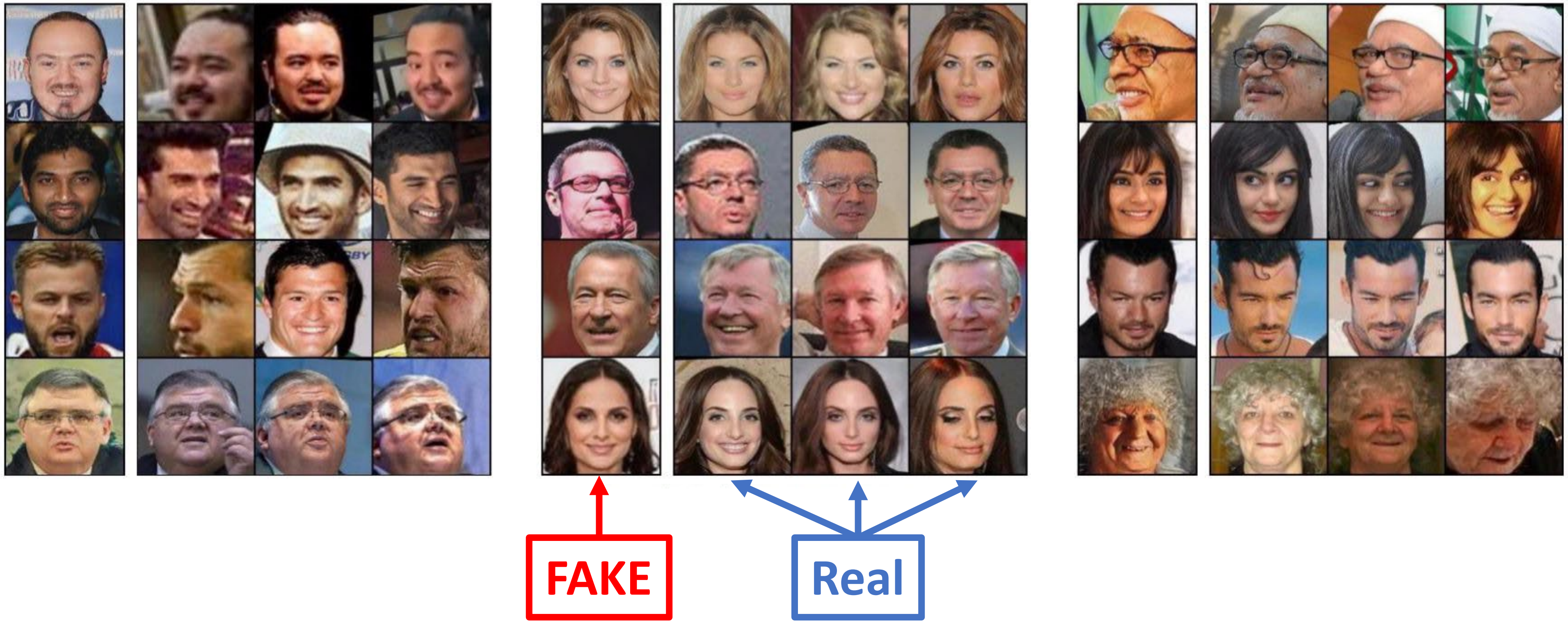
By Will Douglas Heaven

October 12, 2021

https://arxiv.org/abs/2107.06304

Deep Neural Networks are Surprisingly Reversible: A Baseline for Zero-Shot Inversion

Xin Dong[1,2]*, Hongxu Yin[1], Jose M. Alvarez[1], Jan Kautz[1], and Pavlo Molchanov[1]
[1]NVIDIA, [2]Harvard University
xindong@g.harvard.edu, {dannyy, josea, pmolchanov, jkautz}@nvidia.com



When ML Goes “Boink”

- Mimic
  - Insufficient training data can lead to “mimicking” of original records
- Membership Inference*
  - User can test if features of someone they know appear to be in the training data
  - Requires knowing the features in question
- Attribute Inference
  - User can predict features (they don’t know) about someone based on features they do know
- Combining Membership and Attribute is where disclosure occurs

*Zhang, Yan, Malin. Membership inference attacks against synthetic health data. Journal of Biomedical Informatics. 2022

Most Importantly

- We must ensure that there is clinical face value in the data.

- This takes much more time than evaluating the statistical viability

- AI is getting better, but much of medicine still requires human intuition

  (it’s an “open world” problem)

Some Parting Thoughts

- The problems we face are enormously complex and likely beyond our current recognition

- Our current ethics quandaries will take a long time to address

- Ethics should not be addressed *after* AI is created

- Engage. Educate. Evaluate.

Acknowledgements

- Toufeeq Ahmed (Vanderbilt)
- Shilo Anders (Vanderbilt)
- Cinnamon Bloss (UCSD)
- Victor Borza (Vanderbilt)
- Thomas Brown (Vanderbilt)
- Alex Carlisle (NADPH)
- You Chen (Vanderbilt)
- Hoon Cho (Broad)
- Ellen Clayton (Vanderbilt)
- Joseph Coco (Vanderbilt)
- Benjamin Collins (Vanderbilt)
- Carolyn Diehl (Vanderbilt)
- Joyce Harris (Vanderbilt)
- Paul Harris (Vanderbilt)
- Rachele Hendricks-Sturrup (Duke, NADPH)
- Xiaoqian Jiang (UTHSC)
- Murat Kantarcioglu (UT Dallas)
- Yejin Kim (UTHSC)
- Chris Lindsell (Vanderbilt)
- Michael Matheny (Vanderbilt)
- Camille Nebecker (UCSD)
- Laurie Novak (Vanderbilt)
- Lucila Ohno-Machado (UCSD)
- Kirk Roberts (UTHSC)
- Babak Salimi (UCSD)
- Malaika Simmons (NADPH)
- Berk Uston (UCSD)
- Eugene Vorobeychik (WUSTL)
- Zhiyu Wan (Vanderbilt)
- Colin Walsh (Vanderbilt)
- Martin Were (Vanderbilt)
- Chao Yan (Vanderbilt)
- Zhijun Yin (Vanderbilt)
- Xinmeng Zhang (Vanderbilt)
- Ziqi Zhang (Vanderbilt)

Questions? Comments? Discussion?

b.malin@vumc.org

Center for Genetic Privacy and Identity in Community Settings

https://www.vumc.org/getprecise

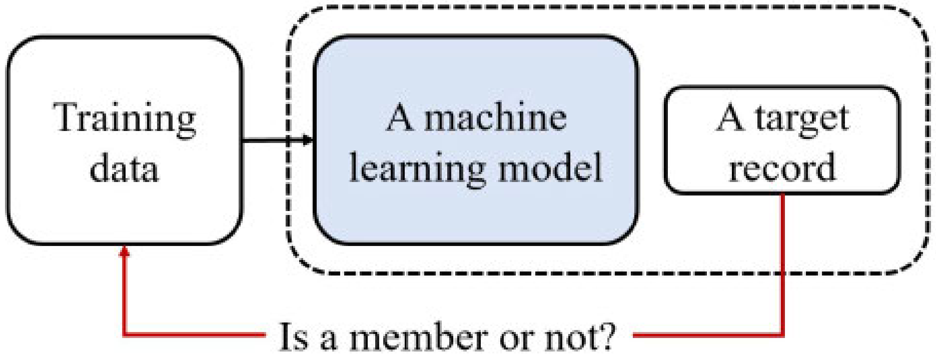
Bridge2AI Ethics and Trustworthy AI Core

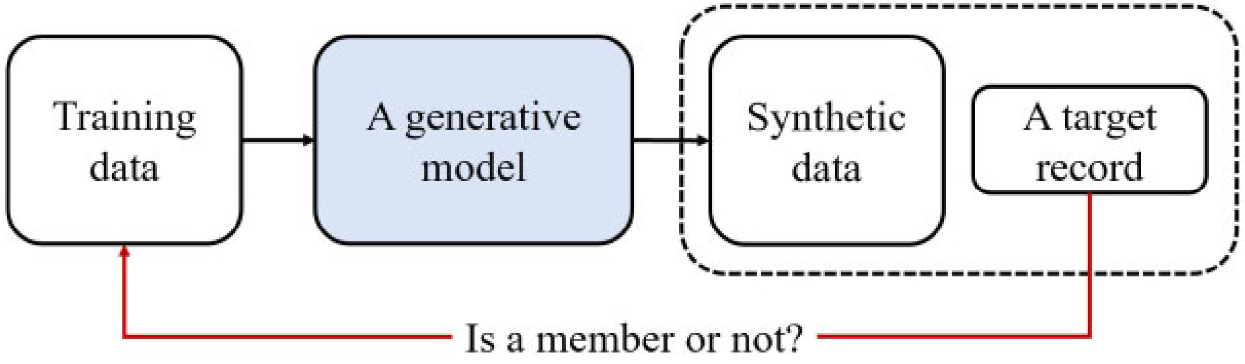
https://bridge2ai.org/ethics-core/

AIM-AHEAD Applied AI Ethics Team

https://aim-ahead.net/

Membership Intrusion

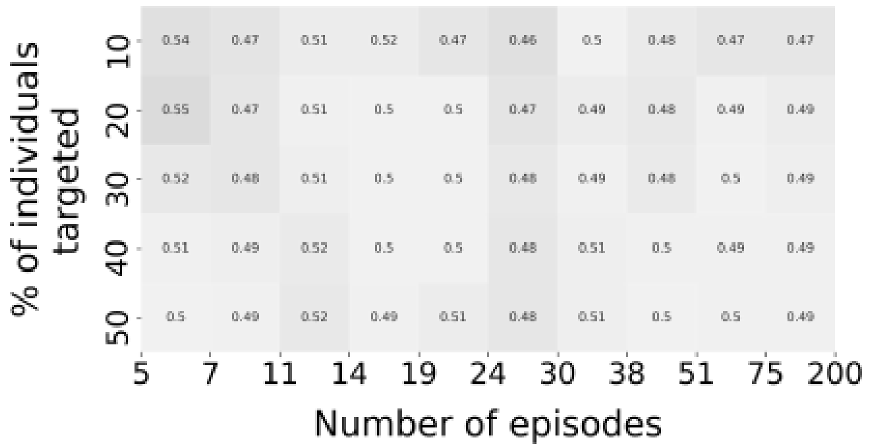




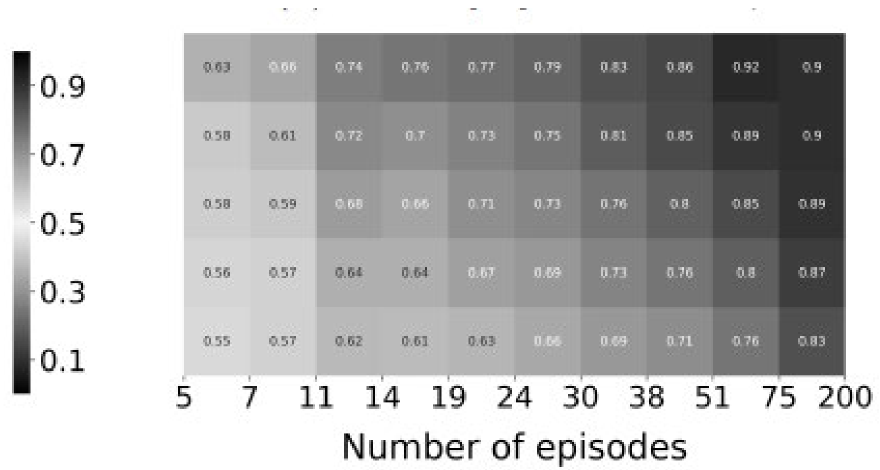
Zhang, Yan, Malin. Journal of Biomedical Informatics. 2022.

An Attack on VUMC Data

- 45,000 patients, diagnosis and procedure codes
- Up to 200 visits
- Adversary has 10% “prior” knowledge



Fully Synthetic



Partially Synthetic