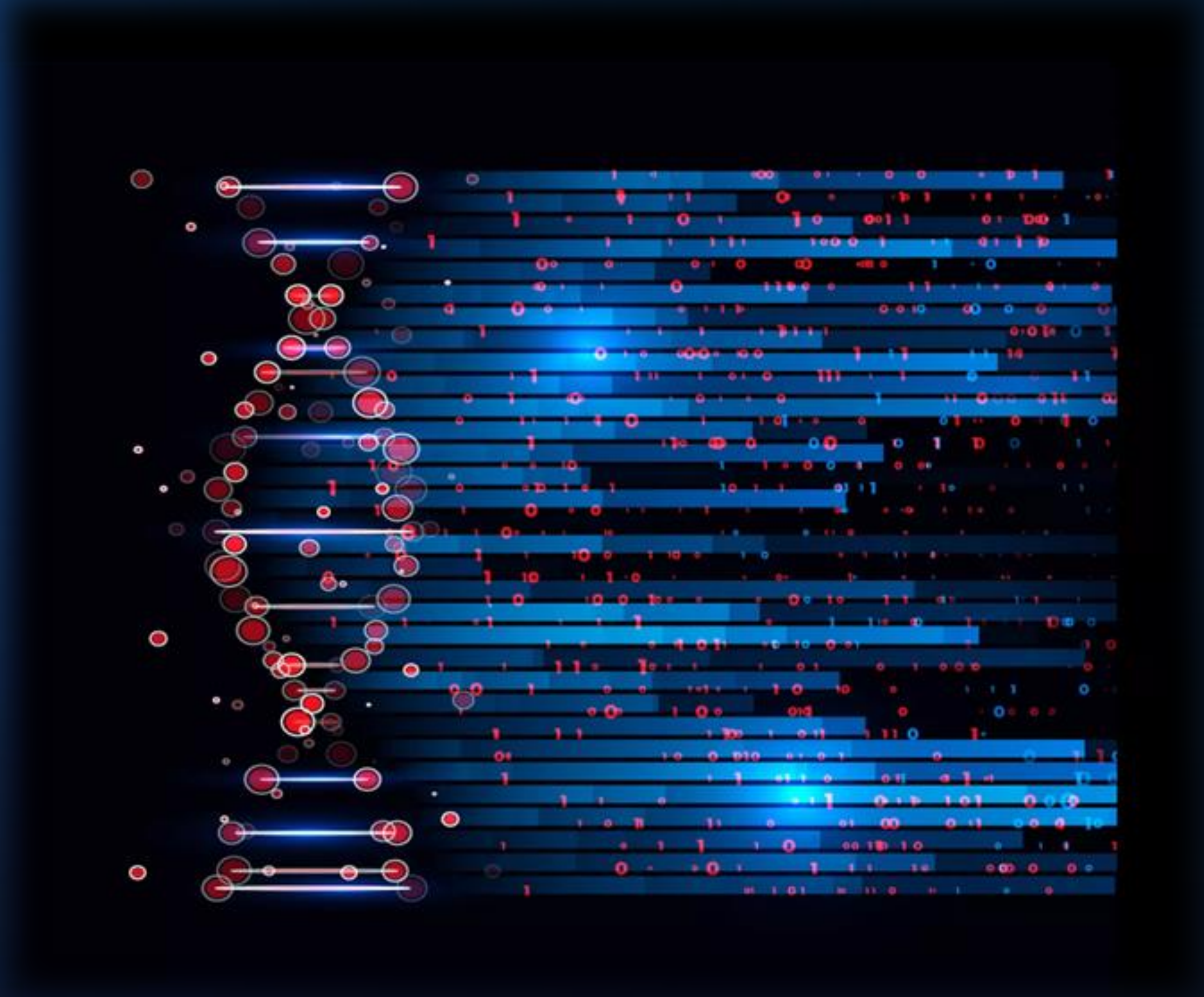
# NIST NCCoE Genomic Data PETs Testbed & Dioptra Webinar



**June 9, 2026**

**1:00 p.m. – 3:30 p.m. EDT**

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

# Agenda

| Time (EDT) | Topic | Speaker |
|---|---|---|
| 1:00 – 1:05 | **Opening Remarks** | • Cherilyn Pascoe, Director, NIST/NCCoE |
| 1:05 – 1:15 | **Webinar Objectives** | • Justin Wagner, Co-Principal Investigator, NIST<br>• Ron Pulivarti, Co-Principal Investigator, NIST/NCCoE |
| 1:15 – 1:20 | **Project Overview and Roadmap** | • Martin Wojtyniak, Project Lead, MITRE/NCCoE |
| 1:20 – 1:50 | **Session 1: Privacy-Enhancing Technologies (PETs)**<br>• Overview of the NIST PETs Testbed and Results | • Gary Howarth, NIST |
| 1:50 – 2:15 | • Results from the Genomics Privacy-Preserving Federated Learning 2025 Red Teaming Event | • Christine Task, Knexus Research |
| 2:15 – 2:40 | • PETs Q&A | • Moderator: Gary Howarth, NIST |
| 2:40 – 3:10 | **Session 2: NIST Dioptra**<br>• Overview of NIST's software test platform for assessing AI Models | • Keith Manville, MITRE/NCCoE |
| 3:10 – 3:25 | • Dioptra Q&A | • Moderator: Keith Manville, MITRE/NCCoE<br>• Moderator: James Glasbrenner, MITRE/NCCoE |
| 3:25 – 3:30 | **Closeout** | • Justin Wagner, Co-Principal Investigator, NIST<br>• Ron Pulivarti, Co-Principal Investigator, NIST/NCCoE |

# NCCoE Welcome
# *Cherilyn Pascoe, NIST*

- Director, National Cybersecurity Center of Excellence

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

# About the NCCoE

The NIST National Cybersecurity Center of Excellence (NCCoE) is a **collaborative hub convening experts from industry, government, and academia** to solve organizations' most pressing cybersecurity challenges.

***Mission:***

*Accelerate the adoption of secure technologies.*

# A U.S. Cybersecurity Innovation Hub

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

## Over 180 collaborative agreements providing technology and expertise to support our work

ami · aws · BedRock Systems Inc · BROADCOM
CISCO · CLOUDFLARE · cloudure · COMCAST
CYBERARK · Delinea · DELL Technologies · digicert
DRAGOS · f5 · FORESCOUT · FORTINET
GLOBAL CYBER ALLIANCE · Google Cloud · HPE · IBM
iddataweb · intel · JPMorgan Chase & Co. · Lookout
MANDIANT · Motorola Solutions · Microsoft · NEXTLABS
nok nok · okta · paloalto networks · SIA Spherical Analytics
tdi technologies · tenable · THALES · Trellix
ZIMPERIUM · zscaler

## The NCCoE's Impact

### Strengthen U.S. Cybersecurity

Provide practical guides to help organizations implement standards-based, repeatable, and scalable solutions

### Drive Standards Adoption & Innovation

Reveal opportunities to improve standards to better address real-world challenges

### Improve Technology

Help vendors strengthen products' security and interoperability by building actual demonstrations

### Foster Public-Private Innovation

Convene cybersecurity experts from industry, academia, and government to develop integrated solutions

# NCCoE Pillars

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

## Data Protection

Cryptography, Identity, and Privacy

## Trusted Enterprise

Foundational Infrastructure & Hardware Roots of Trust

## Cybersecurity & Artificial Intelligence

Cybersecurity of AI Tools & AI for Cybersecurity

## Resilient Embedded Systems

Cybersecurity of OT/ICS & IoT

# How to Get Involved

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

### National Cybersecurity Excellence Partnership (NCEP)

Partners provide direction, resources and expertise to support the NCCoE's work

### Cooperative Research and Development Agreement (CRADA)

Collaborators who provide hardware, software, and/or expertise for a project

### Community of Interest (COI) Members

Individuals who want to learn about and provide input on the NCCoE's work

### Interagency Agreements (IAAs)

Federal agencies who sponsor research projects at the NCCoE

### Subscribe for Updates

Receive updates on NCCoE projects and publications

### Join Events

Participate in the NCCoE's events and discussions

### Provide Feedback

Review and provide feedback on our work

# NIST/NCCoE Presenters

**Ron Pulivarti**
NIST
National Cybersecurity Center of Excellence

**Justin Wagner**
NIST
Material Measurement Laboratory

**Martin Wojtyniak**
MITRE/NCCoE

**Gary Howarth**
NIST
Privacy Engineering Program

**Keith Manville**
MITRE/NCCoE

# Webinar Objectives

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

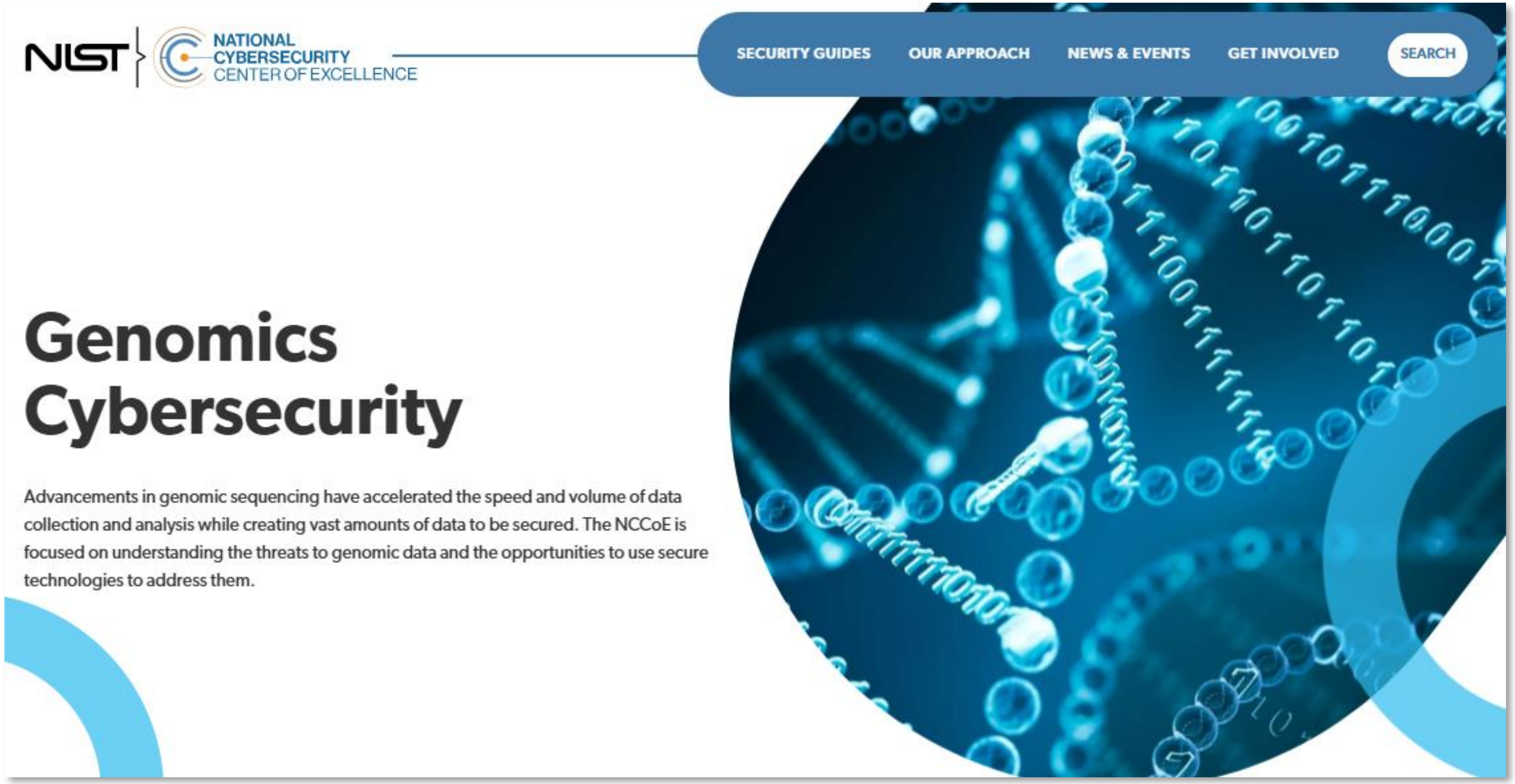
- Genomic data has high economic value for biotechnology development but is among the most sensitive data types at a personal level (NIST IR 8432 *Cybersecurity of Genomic Data*). At the same time, AI is increasing the value and use of genomic data.

- NIST is addressing genomic data cybersecurity and privacy risks through:
    - Threat modeling -- draft publications and webinar/workshop recordings
    - PETs testbed -- data and evaluation metrics
    - AI evaluation infrastructure (Dioptra) -- test platform

- How you can help collaborate with us:
    - Provide feedback during the webinar w/ our Q&A sessions and Slido
    - Email project teams (genomic_cybersecurity_nccoe@nist.gov, pets@nist.gov, dioptra@nist.gov)
    - Submit comments to draft publications as they are released

# Project Overview and Roadmap

# NCCoE Genomics Cybersecurity

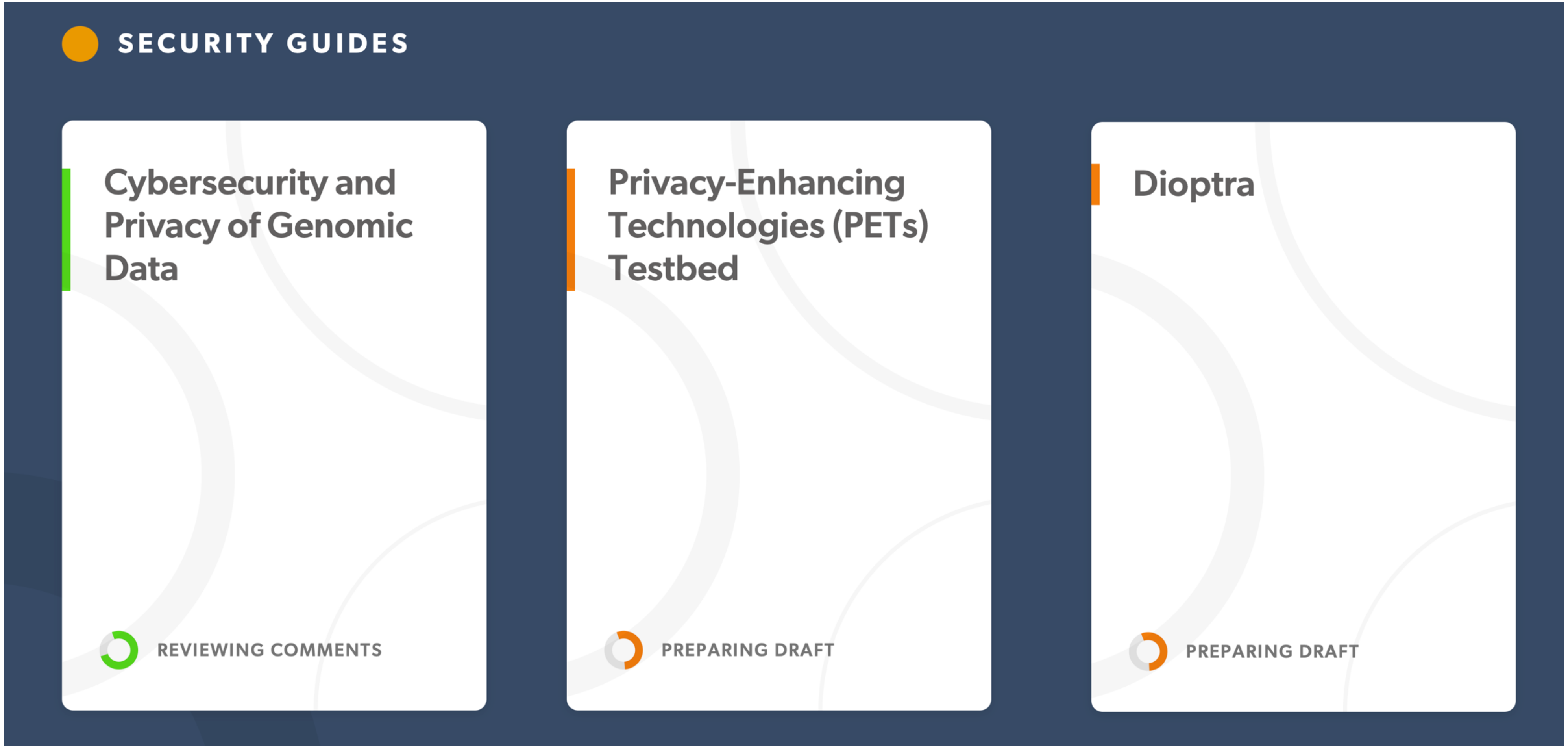



1. **Join our Community of Interest (COI)**
2. **Stream Workshops & Webinars**
3. **Review Publications & other Resources**

https://www.nccoe.nist.gov/genomics-cybersecurity

# NCCoE Genomics Cybersecurity






https://www.nccoe.nist.gov/genomics-cybersecurity

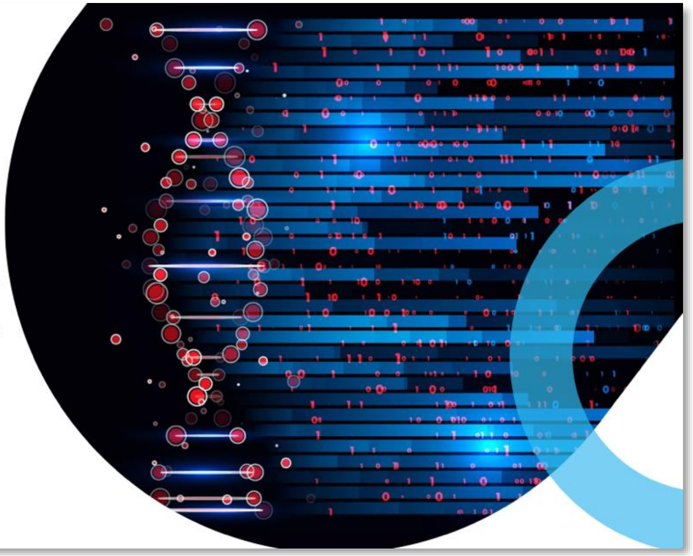
# NCCoE Genomic Data Profile and Threat Modeling

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

## Cybersecurity and Privacy of Genomic Data

The advent of innovative genomic sequencing technologies has ushered in an era where it is now possible to sequence and analyze an entire genome quickly and affordably. While vast amounts of genomic data have helped fuel our nation's economic, technological, and health leadership posture, this information may not be sufficiently protected to prevent misuse.

READ THE 2-PAGE FACT SHEET



https://www.nccoe.nist.gov/genomics-cybersecurity

**STATUS: REVIEWING COMMENTS**

The public comment period has closed for NIST Special Publication 1800-43, *Genomic Data Threat Modeling: An Implementation for Genomic Data Sequencing and Analysis*.

- NIST SP 1800-43A: Executive Summary (Draft)
- NIST SP 1800-43C: Privacy (Draft)
- Genomic Data Threat Modeling NIST Pages Static Site

---

The NCCoE Genomic Data team has finalized NIST IR 8432, *Cybersecurity for Genomic Data*. The public comment period has closed for NIST IR 8467, *Genomic Data Cybersecurity and Privacy Frameworks Community Profile* and NIST CSWP 35, *Cybersecurity Threat Modeling the Genomic Data Sequencing Workflow*.

- NIST IR 8432 Cybersecurity of Genomic Data
- NIST CSWP 35 ipd, Cybersecurity Threat Modeling the Genomic Data Sequencing Workflow (Initial Public Draft)
- NIST IR 8467 2pd, Genomic Data Cybersecurity and Privacy Frameworks Community Profile (Second Public Draft)
- NIST IR 8467 Genomic data profile spreadsheet

# Project Roadmap

NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

2021 | 2022 | 2023 | 2024 | 2025 | 2026

## CHALLENGES

Identified and engaged SMEs
***Genomic Data Community of Interest (COI)***

***NIST IR 8432***
*Cybersecurity of Genomic Data*
***Final 2023***

Conducted workshops highlighting challenges and solutions for protecting genomic data
***Workshops & Subsequent Webinars***

## CYBERSECURITY & PRIVACY

***NIST IR 8467***
*Genomic Data CSF & PF Community Profile (2pd)*
***December 2024***

***Coming Soon:***
Update Profile to PF 1.1

***NIST Privacy-Enhancing Technologies (PETs) Testbed***
Red Teaming Event
***2025***
Testbed in development

***NIST SP 1800-43***
*Threat Modeling for Genomic Data:*
*An Implementation for Genomic Sequencing and Data Analysis*

## AI

***NIST Privacy-Enhancing Technologies (PETs) Testbed***
Platform Development

***NIST Dioptra***
Platform Development and testing with genomic models

# NCCoE PETs Testbed

**Gary Howarth**

Physical Scientist & Privacy Engineering Program Manager
NIST

# Why does NIST research privacy?

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

15 U.S.C. § 272c

(19) facilitate and support the development of a voluntary, consensus-based set of technical standards, guidelines, best practices, methodologies, procedures, and processes to improve privacy protections in systems, technologies, and processes used by both the public and private sector;

(20) support privacy measures, including voluntary, consensus-based technical standards, best practices, guidelines, metrology, and testbeds for the design, adoption, and deployment of privacy enhancing technologies;

# How does NIST research privacy?

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Risk Management Tools

- Privacy Framework

Privacy Technology Benchmarking

- Collaborative Research Cycle (synthetic data)
  - Benchmark datasets
  - Fidelity, utility, and privacy metrics (SDNist)
  - Privacy Red-Teaming
- Privacy-Preserving Federated Learning of Genomic Data
  - Model PPFL architecture (coming this summer)
  - Datasets (coming this summer)
  - Privacy Red-Teaming

# Difficulties with protecting privacy

# Protecting data release with redaction?

| Record Number | Name | DOB | Sex | Address | Date of Visit | Reason for Visit |
|---|---|---|---|---|---|---|
| 132313 | John Doe | 2/2/1979 | Male | 32 Water St.<br>Northampton, MA 01129 | 9/5/1997 | Suicide attempt |
| 318977 | Clara Vasquez | 5/19/1992 | Female | 742 Evergreen Terrace<br>Springfield, MA 01020 | 9/5/1997 | Lead poisoning |
| 218987 | Julio Vasquez | 6/12/1994 | Female | 742 Evergreen Terrace<br>Springfield, MA 01020 | 9/5/1997 | Lead poisoning |
| 156465 | William Wild | 7/31/1949 | Male | 312 Harvey St.<br>Cambridge, MA 03129 | 9/5/1997 | Back pain |

# Protecting data release with redaction?

| Record Number | Name | DOB | Sex | Address | Date of Visit | Reason for Visit |
|---|---|---|---|---|---|---|
| 132313 | [redacted] | [redacted]1979 | Male | [redacted] Northampton, MA 01129 | [redacted]/1997 | Suicide attempt |
| 318977 | [redacted] | [redacted]1992 | Female | [redacted] Springfield, MA 01020 | [redacted]/1997 | Lead poisoning |
| 218987 | [redacted] | [redacted]1994 | Female | [redacted] Springfield, MA 01020 | [redacted]/1997 | Lead poisoning |
| 156465 | [redacted] | [redacted]1949 | Male | [redacted] Cambridge, MA 03129 | [redacted]/1997 | Back pain |

# Can you protect PII with redaction?

Redacted data is vulnerable to *de-anonymization* attacks with auxiliary data sources

**Redacted Medical Record**

- Diagnosis
- Procedure
- Medications
- Year of visit
- Ethnicity
- Zip code
- Birth year
- Sex

**Public Voter Roster**

- Name
- Party affiliation
- Address
- Registration date
- Date last voted
- Zip code
- Birth year
- Sex

# Protecting PII with Redaction?



Redacted data is vulnerable to *de-anonymization* attacks with auxiliary data sources

Redacted Medical Record

- Diagnosis
- Procedure
- Medications
- Date of Visit
- Ethnicity

- Zip code
- Birth year
- Sex

Public Voter Roster

- Name
- Party affiliation
- Address
- Registration date
- Date last voted

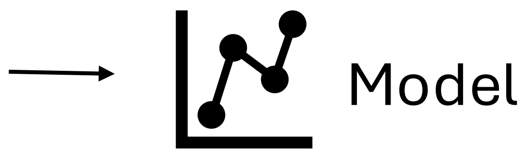
87% of people in U.S. can be re-identified using 3 quasi-identifiers.

L. Sweeney. Weaving Technology and Policy Together to Maintain Confidentiality. *Journal of Law, Medicine & Ethics*, 25, nos. 2&3 (1997): 98-110.

L. Sweeney. *Simple Demographics Often Identify People Uniquely*. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

# Machine Learning on Genomes

| Person | Lactose | DNA Sequence | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Bob | tolerant | G | C | G | G | C | C | ... |
| Alice | tolerant | A | C | G | G | C | C | ... |
| Greg | intolerant | A | C | G | C | T | C | ... |
| Lisa | Intolerant | A | C | G | G | T | C | ... |

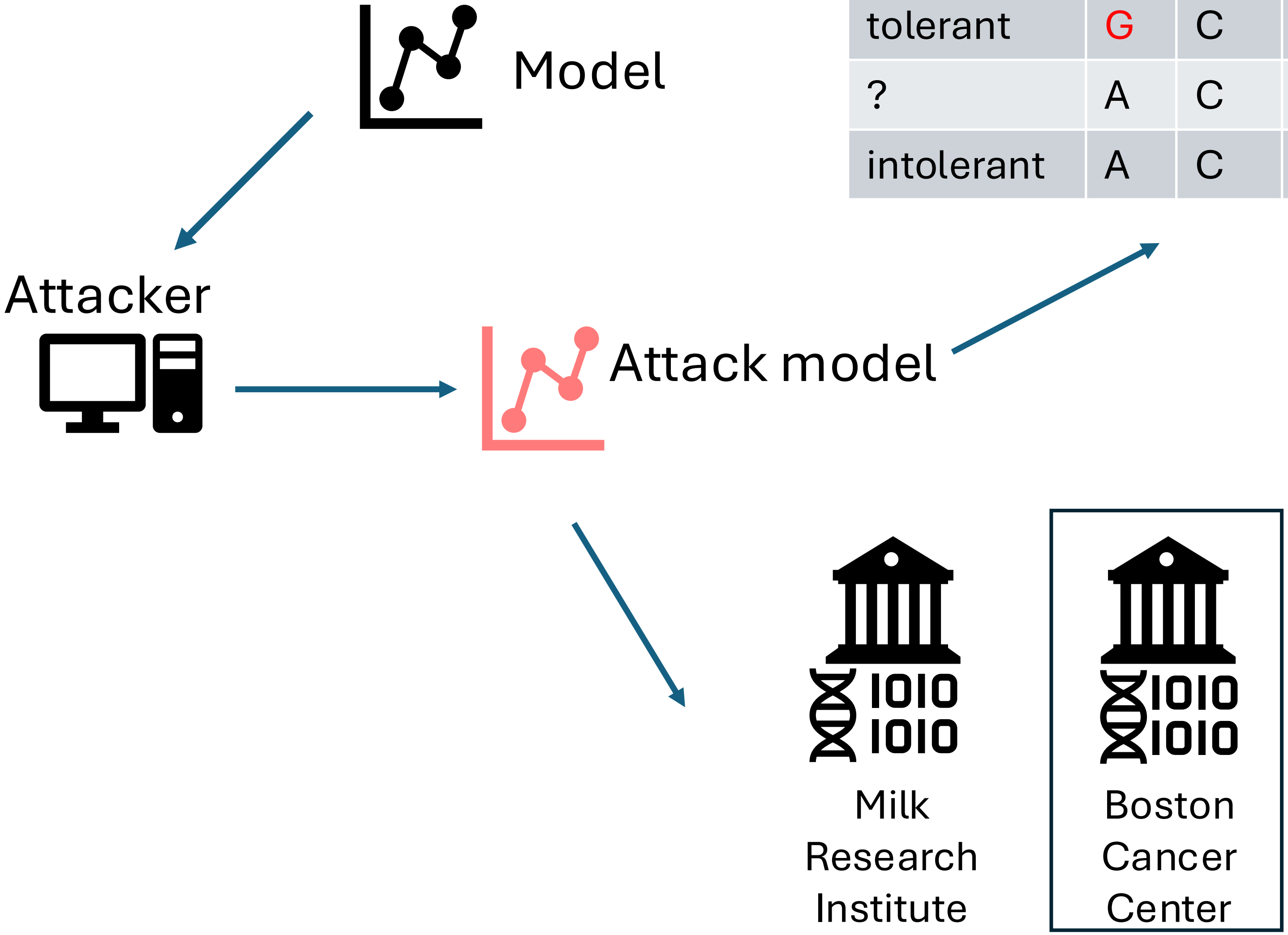



| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Gary** | **?** | **A** | **C** | **G** | **G** | **T** | **C** | **...** |

Model predicts likelihood of Gary being lactose tolerant based on his DNA.

# Machine learning models are vulnerable to privacy attacks

| Lactose | DNA Sequence | | | | | |
|---|---|---|---|---|---|---|
| tolerant | G | C | G | G | C | C | ... |
| ? | A | C | ? | G | C | C | ... |
| intolerant | A | C | G | C | T | ? | ... |



# Privacy Enhancing Technologies (PETs)

Technical tools, controls, and frameworks to reduce privacy risks that arise from collecting, sharing, and analyzing data.

# What are PETs?

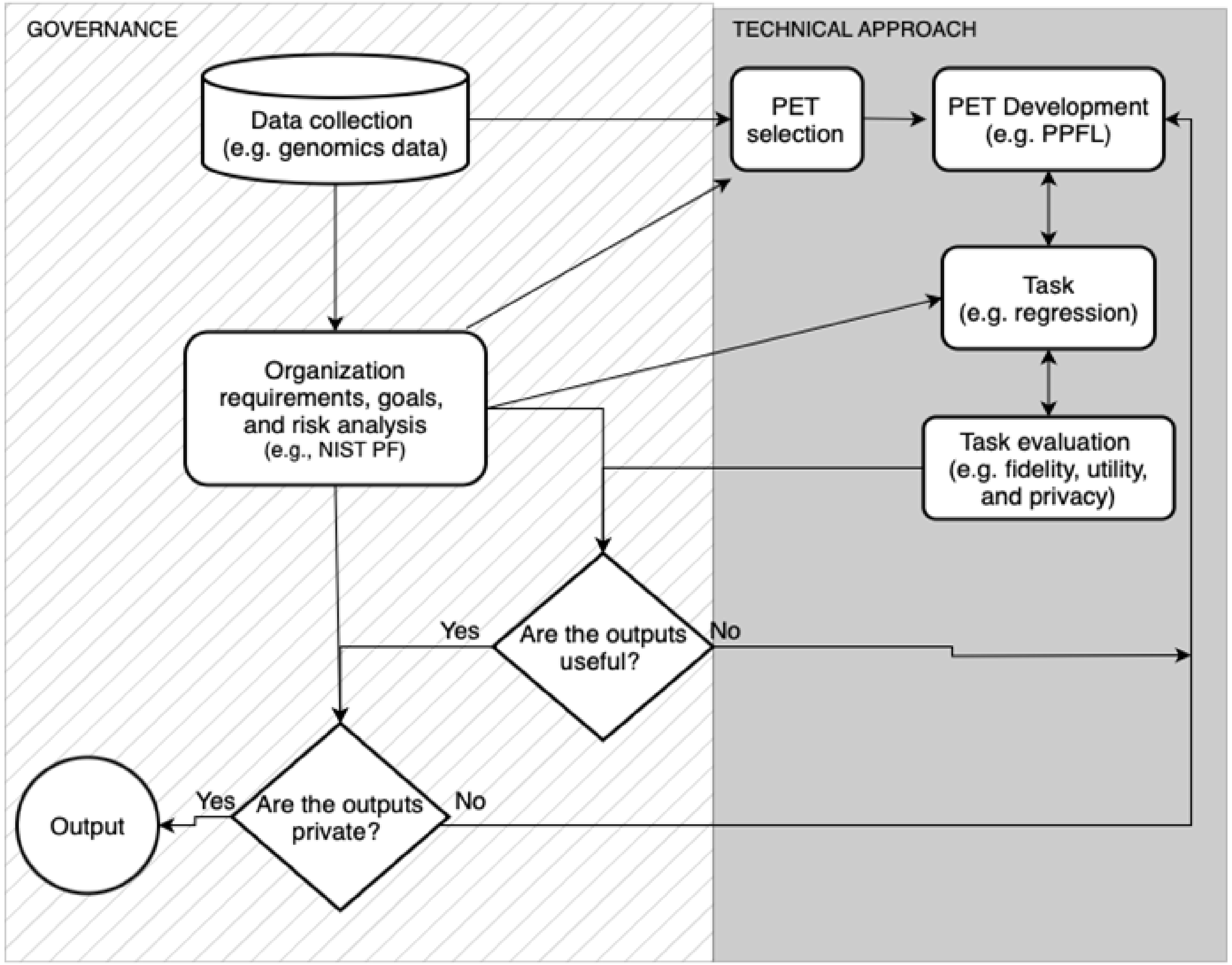
**Privacy Enhancing Technologies (PETs):**

Technical tools, controls, and frameworks to reduce privacy risks that arise from collecting, sharing, and analyzing data.

**Examples:**

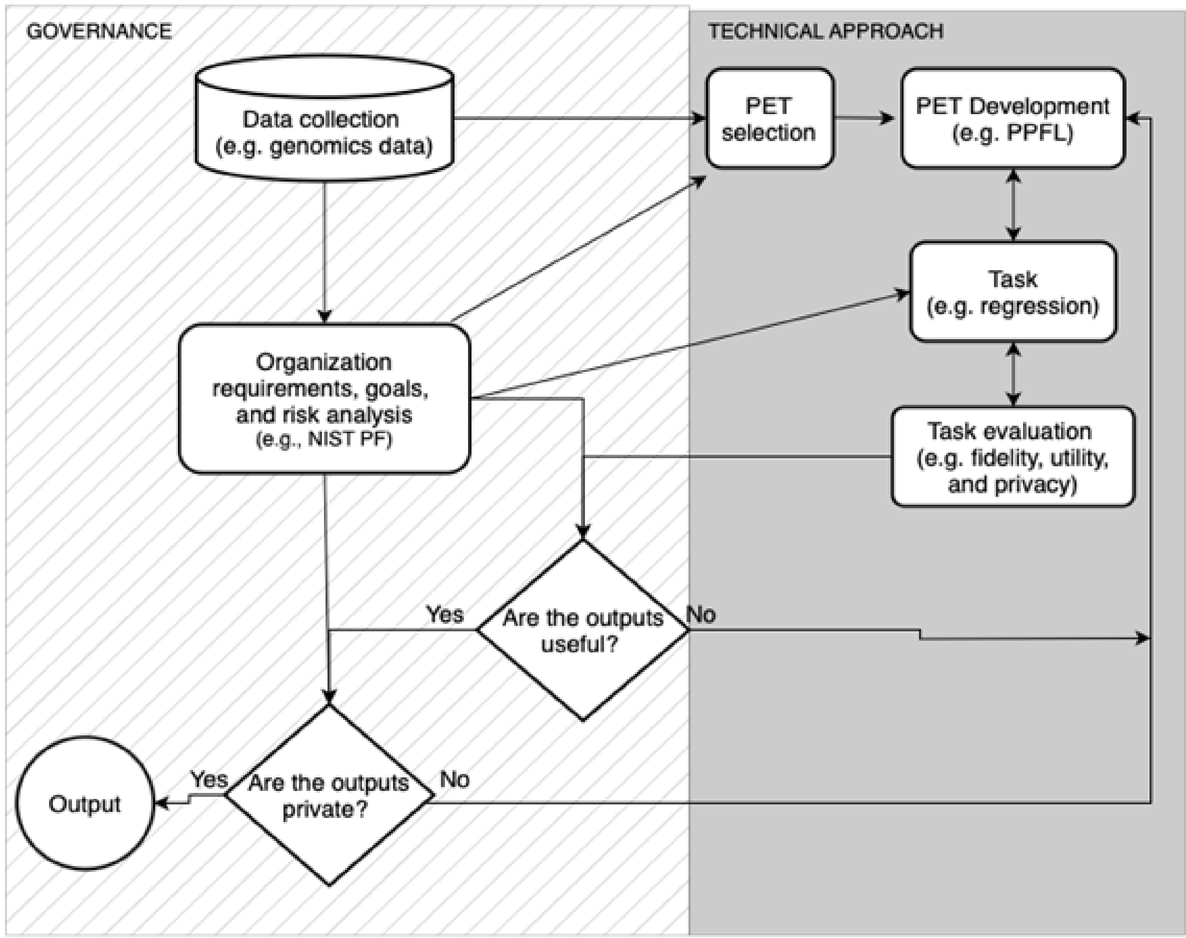
- Encrypting partner data during processing (trusted execution environments / homomorphic encryption)
- Proving you're over 18 without revealing your age (zero knowledge proofs)
- Adding noise to a machine learning model to prevent overfitting (differential privacy)

# Steps to PETS Deployment



# Barriers to PETs Deployments

- Organizations struggle with *every* step of this process
- We lack a common language to talk about these processes
- Very few case studies chronicling this process
- Most organizations have poorly defined data goals
- Every organization believes it's problems and goals are unique
  - PETs deployments tend to be single use only



# Privacy Enhancing Technologies (PETs) Testbed

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

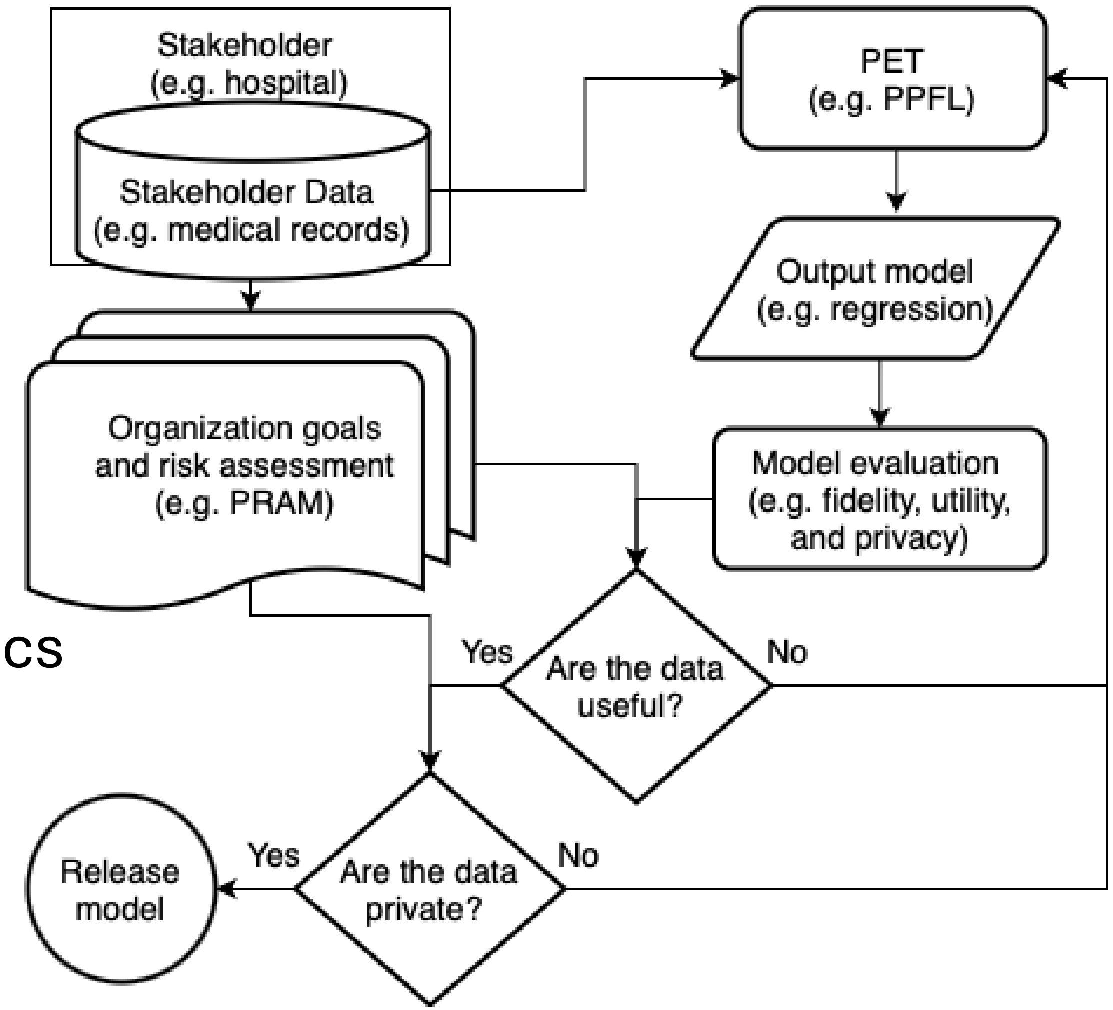
## OBJECTIVE

- Develop model deployments of PETs
- Release benchmark data
- Develop general-use fidelity, utility, and privacy metrology
- Publish privacy and cybersecurity threat modeling and sample mitigations

## SCENARIOS

- Privacy-preserving federated learning of genomics
- Synthetic demographic data
- *Future use cases:* financial transactions, graph data, multi-modal data

NIST PETs Testbed Framework



# Working toward a comprehensive PETs evaluation framework



## Assess

### Data Goals:

- Deriving value from data
- Value to customers
- Business goals

### Constraints:

- Risk assessment
- Legal concerns
- Ethics and values

## Build

### What types of privacy?:

- Input
- Output
- Model
- Linking
- Query

### What PET(s)?

- Find a set of technologies that achieve the mission
- What parameters?

## Evaluate

### Are the data private?

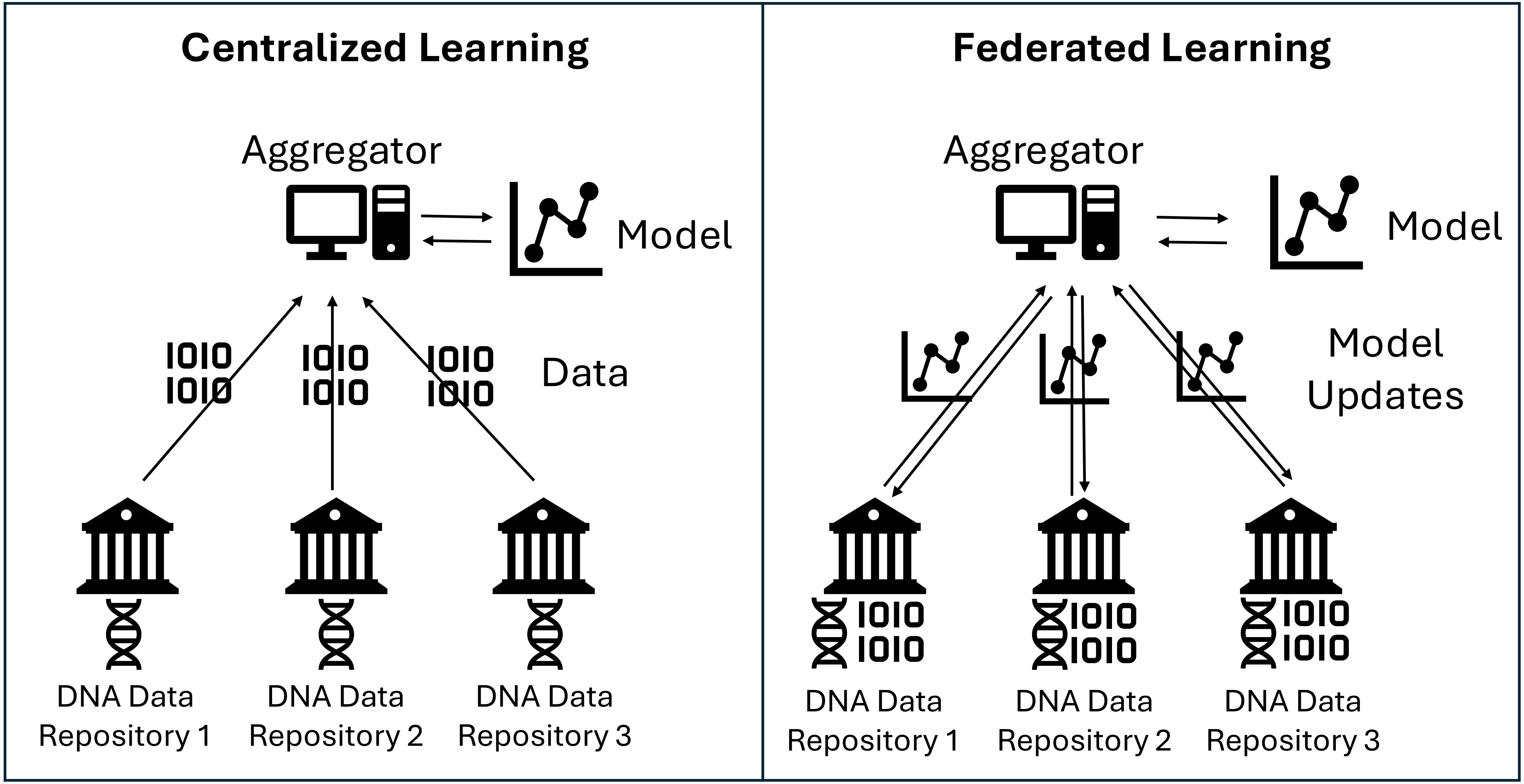
- Theoretical guarantees
- Empirical measurement
- Red teaming

### Are the data useful?

- Benchmark tasks
- Goal achievement evaluation
- Empirical measurement

# Privacy preserving federated learning of genomic data

# Machine learning across databases



# Federated Learning

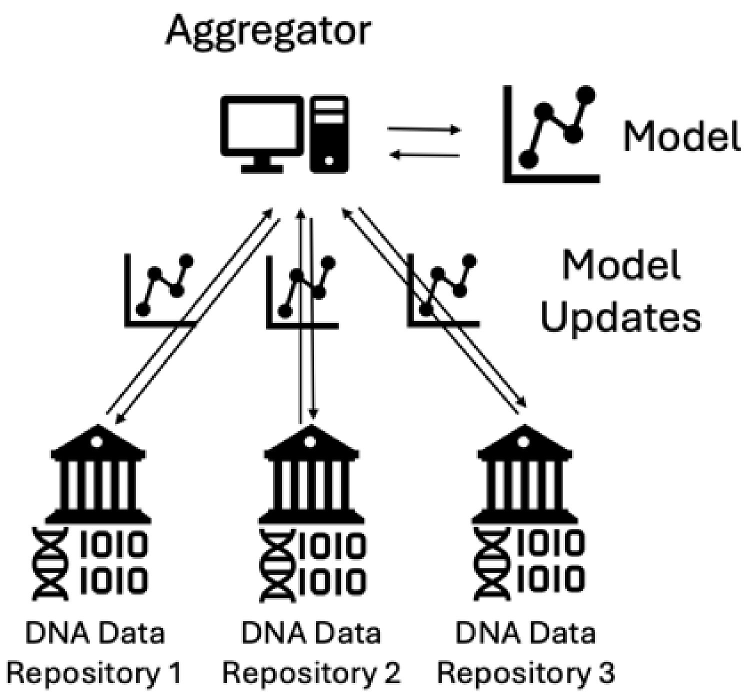
- Federated learning is machine learning on decentralized data
- Training data remains with individual clients

## Federated Learning Process

1. Aggregation server distributes initial model
2. Clients train on local data and send updates back
3. Aggregation server updates model and redistributes
4. Steps 2-3 are repeated until the model converges

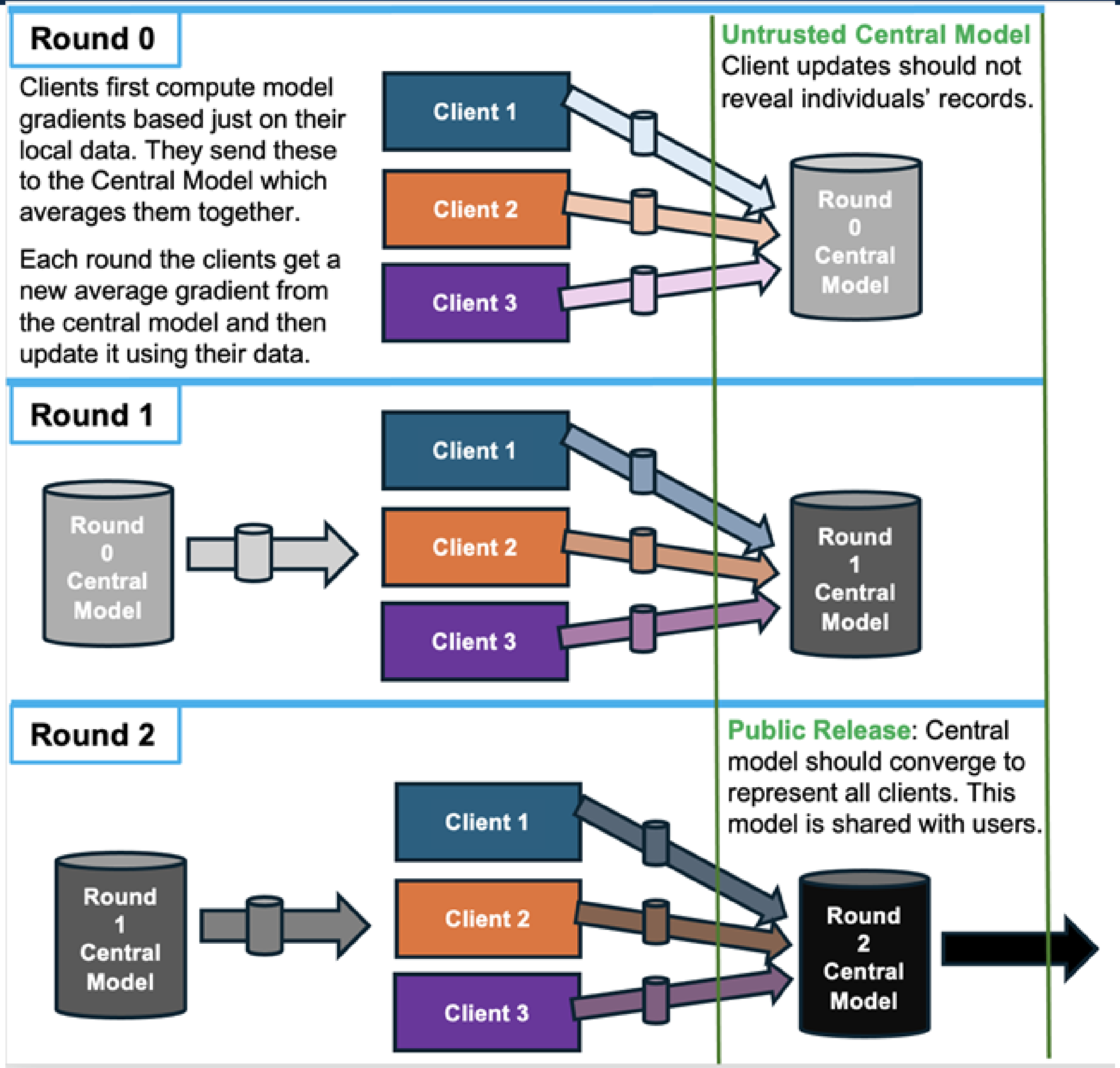
## Example Questions

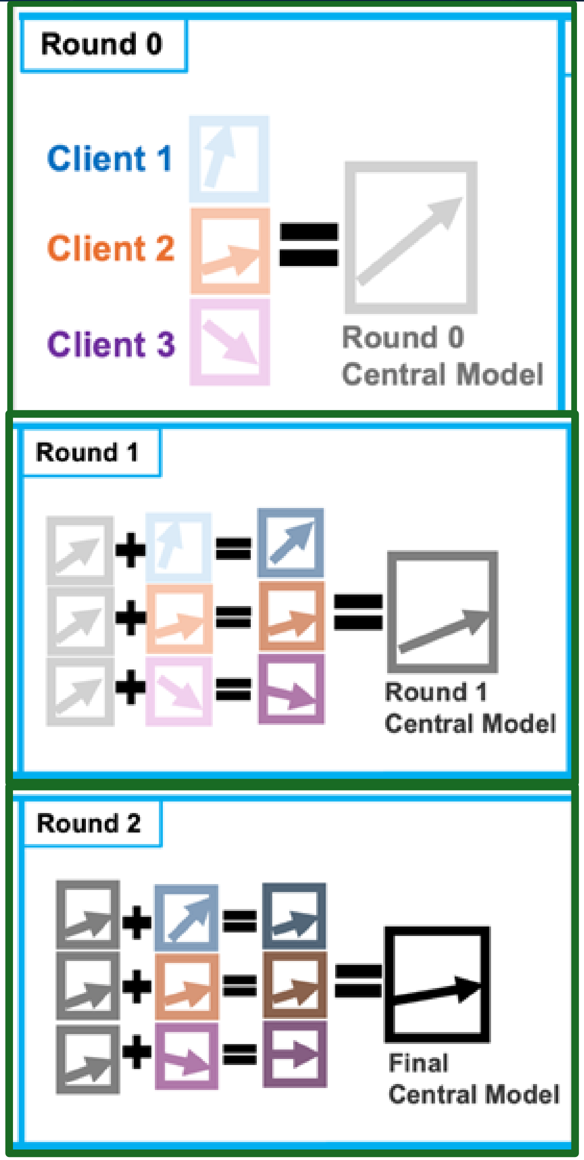
- Which genes contribute to a physical trait (e.g., height)?
- Which mutations are correlated with a specific cancer?



Federated Learning

# Federated Learning Overview



# The Details: Sharing Models

Model Privacy

## What you have:

- An AI model that performs a task
- It was trained on a population of individuals

## What you want to do with it:

- Share it with researchers
- Deploy it in clinics/hospitals
- Combine it with models for the same task

## Privacy risks:

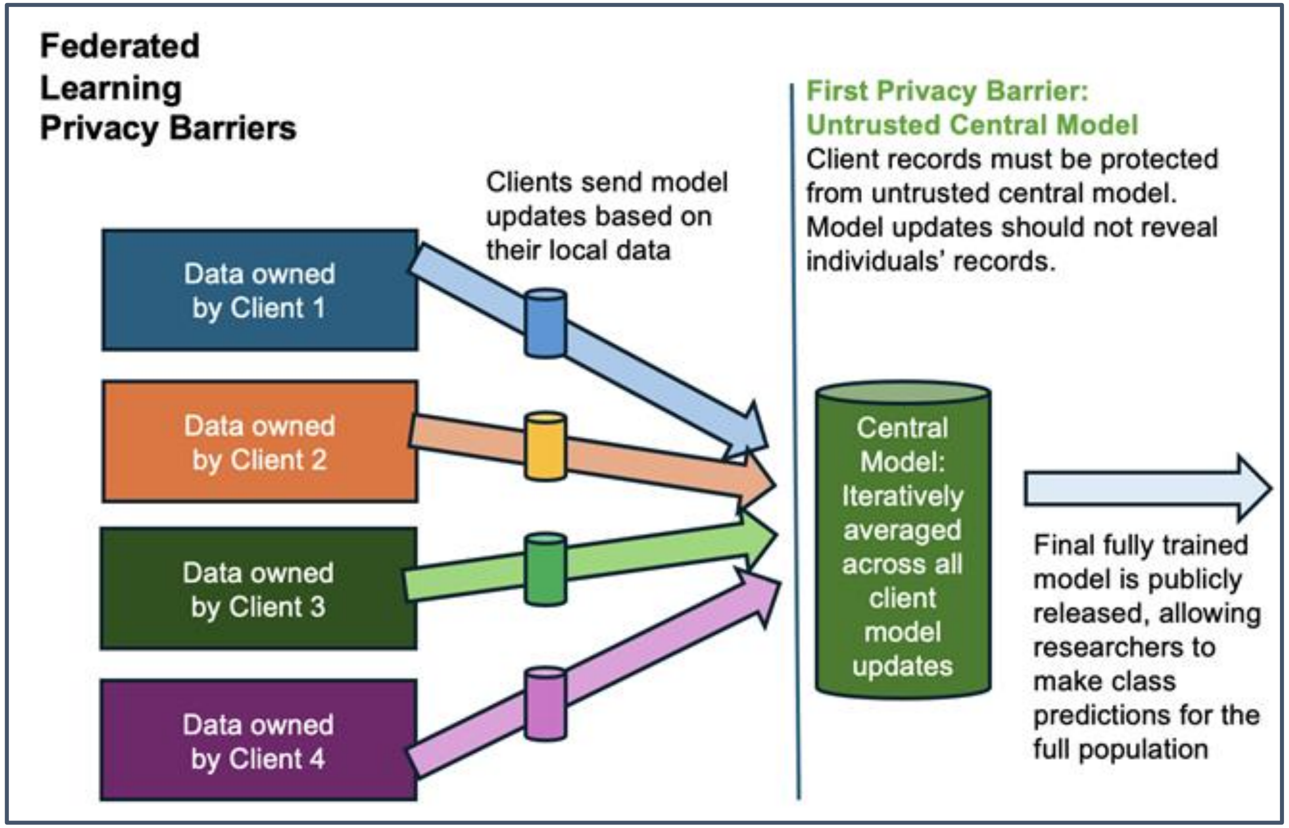
- Membership Attacks, Reconstruction

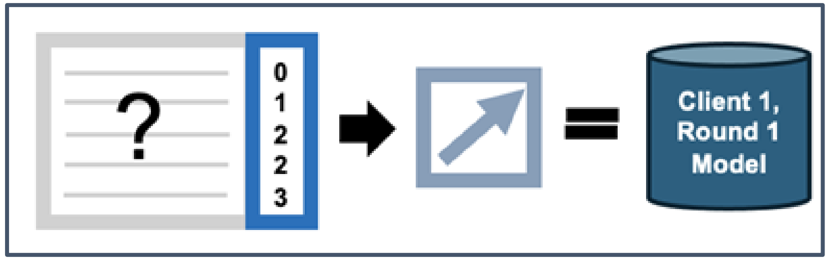
## PETs protection:

- Train with Differential Privacy noise
- Avoid overfit models (larger batch size)
- Train on synthetic data

## What makes it hard:

- Small population, heterogeneous/diverse data
- Outliers important



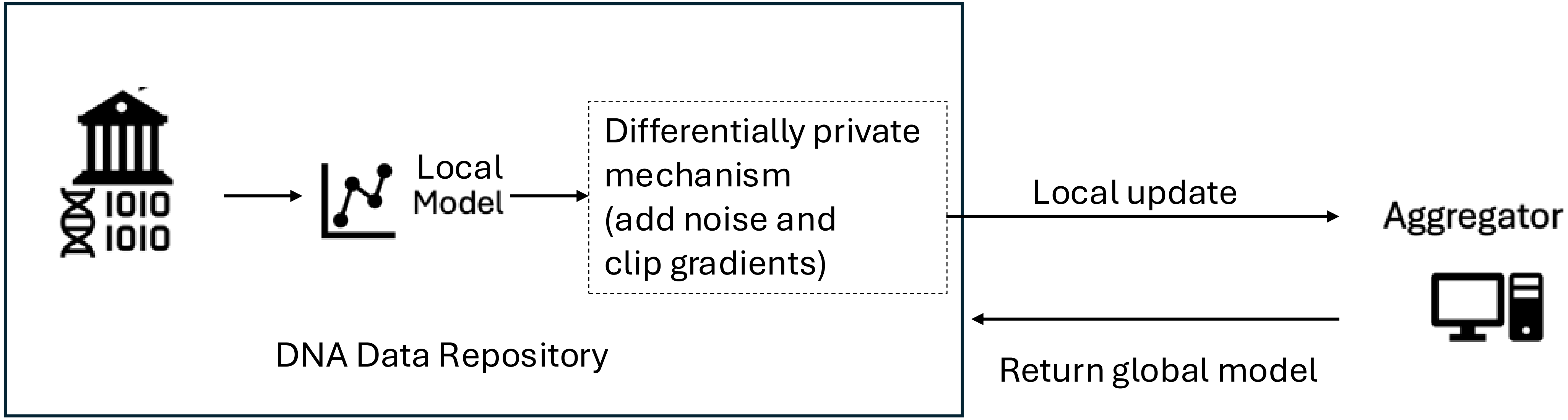

Attackers can sometimes turn model gradients backwards to reconstruct the training data that produced them.

# Potential Vulnerabilities in Federated Learning

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

- **Membership Attacks:** An attacker can use the shared model weights to infer sensitive information about the training data.
- **Data Reconstruction Attacks:** An attacker can reconstruct parts of the training data from the model updates.
- **Model Inversion Attacks:** An attacker can use the shared model weights to infer sensitive information about the training data.
- **Model Evasion and Poisoning Attacks:** An attacker can manipulate the model updates to compromise the model's performance or inject backdoors.

# Privacy-Preserving Federated Learning




**Key Features**

- DP has provable disassociability guarantees
- DP privacy budget (noise) can be adjusted and tuned

**Key Questions**

- How does DP affect global model performance?
- How much protection does DP provide in practice?

# PPFL Testbed

- Created benchmark genomic datasets for machine learning (soybeans and canines)
- Coded a model differentially private federated learning platform
- Implemented Utility and empirical privacy evaluation metrics
- Conducting privacy and cybersecurity risk assessments on system

# PPFL Red Teaming Problems

- Membership Inference Attack
  - Can an attacker predict if a sequence of DNA is also in a data silo?
- Reconstruction Attack
  - Can an attacker learn if individual traits are in batch update?

Full Problem Statements Available at https://pages.nist.gov/genomics_ppfl/pages/problem.html

# PPFL Red Teaming

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

**March – June: Soybean Data**

**Basic Problem (Seed Coat Columns): Membership Inference**

- March 10 Launch
- April 11st First Deadline
- May 1st Second Deadline
- June 2nd Third Deadline
- July 1st Final Deadline

**Advanced Problem: Reconstruction**

- May 2nd Launch
- June 2nd First Deadline
- July 1st Final Deadline

**July: Dog Data**

**Basic Problem: Membership**

- June 30 Problem Launch
- July 29th Final Leaderboard Deadline

**Advanced Problem: Reconstruction**

- June 30 Problem Launch
- July 29th Final Leaderboard Deadline

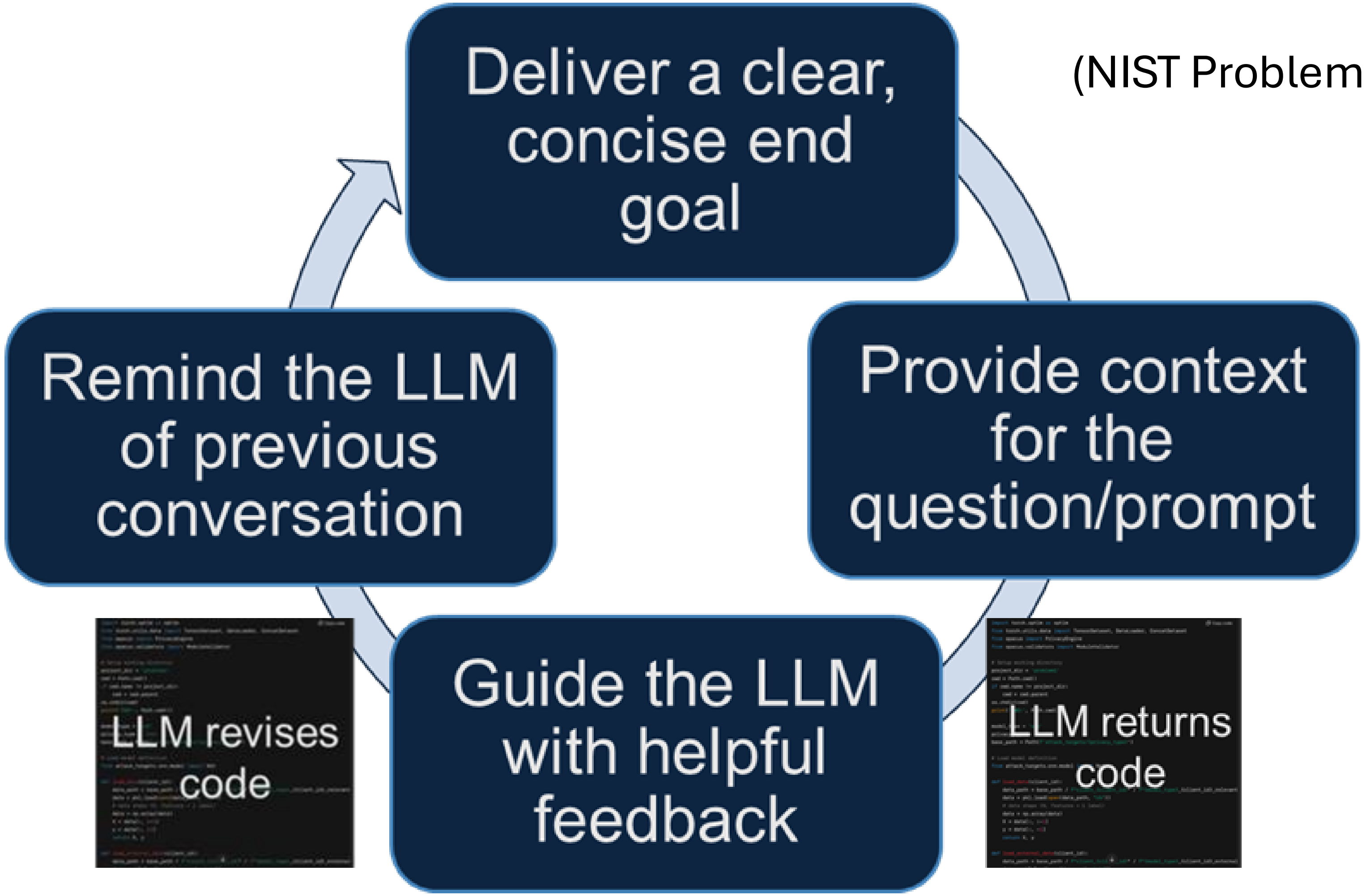
Introduction Webinar Recording -- https://www.nist.gov/video/nist-2025-privacy-red-teaming-introduction

# “m00c0w”’s AI-supported submission

Utilized default OpenAI GPT-4o to create a  highly successful attack model

```python
# ---- Define Model ----
class SimpleClassifier(nn.Module):
    def __init__(self, input_dim, num_classes):
        super().__init__()
        self.fc1 = nn.Linear(input_dim, 64)
        self.bn1 = nn.BatchNorm1d(64)
        self.fc2 = nn.Linear(64, 32)
        self.bn2 = nn.BatchNorm1d(32)
        self.fc3 = nn.Linear(32, num_classes)

    def forward(self, x):
        x = F.relu(self.bn1(self.fc1(x)))
        x = F.relu(self.bn2(self.fc2(x)))
        return self.fc3(x)

input_dim = all_X.shape[1]
num_classes = 5  # 0 = non-member, 1-4 = clients
model = SimpleClassifier(input_dim, num_classes)

# ---- Make model compatible with Opacus ----
model = ModuleValidator.fix(model)
if hasattr(model, 'bn1'):
    model.bn1 = nn.Identity()
if hasattr(model, 'bn2'):
    model.bn2 = nn.Identity()

device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
model = model.to(device)

# ---- Optimizer and DP Engine ----
criterion = nn.CrossEntropyLoss()
optimizer = torch.optim.Adam(model.parameters(), lr=0.003)

privacy_engine = PrivacyEngine()
model, optimizer, train_loader = privacy_engine.make_private(
    module=model,
    optimizer=optimizer,
    data_loader=train_loader,
    noise_multiplier=1.0,
    max_grad_norm=1.0,
)
```

# “m00c0w” AI Workflow: Overview



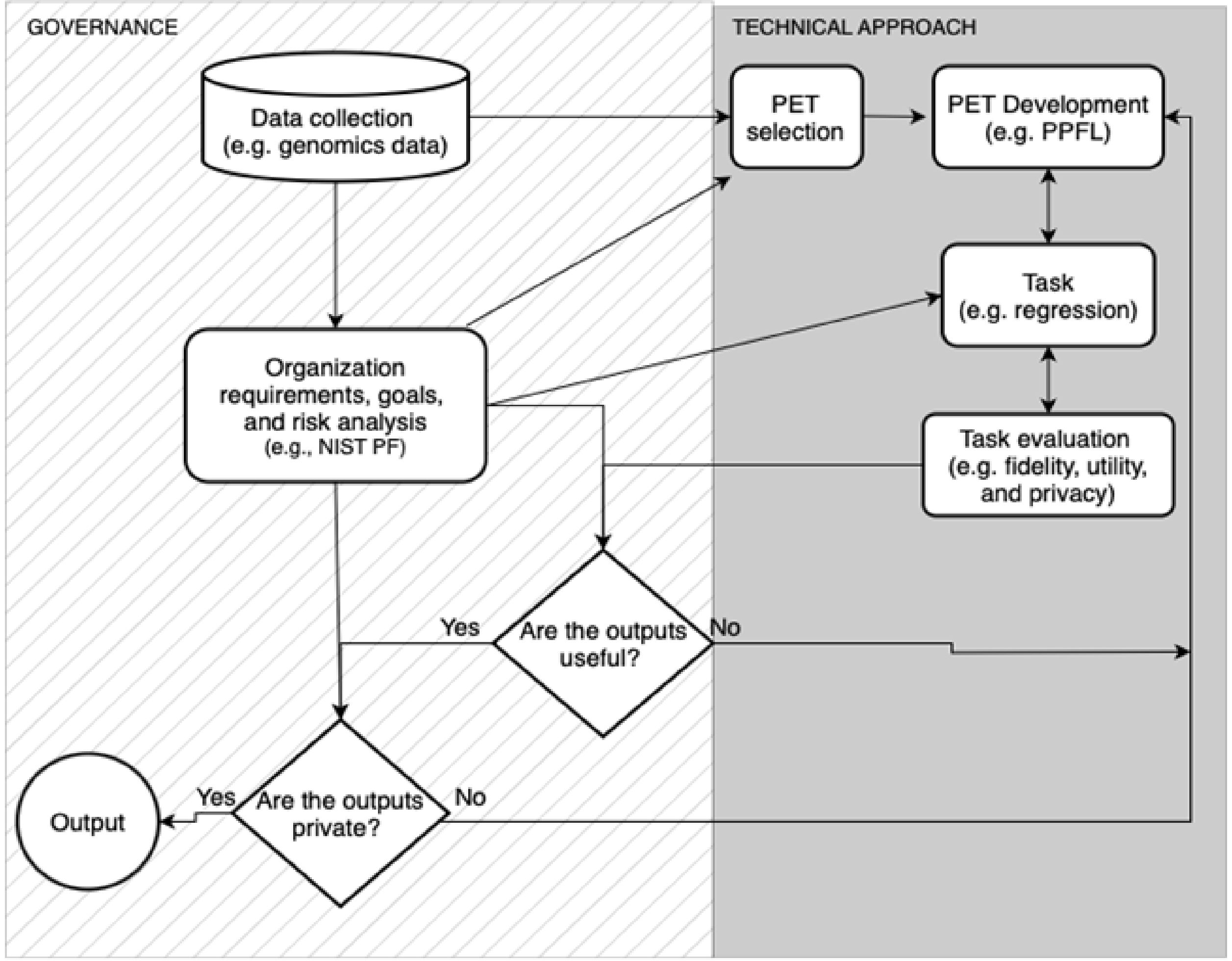

(NIST Problem Statement)

(NIST background on the challenge, attacks, and open-source libraries)

# PETs Testbed

PETs Testbed: (opportunities for ***you*** to get involved)

- Real(istic) problems
- Benchmark, public data
- Model solutions
- Evaluation metrics and tools



# Privacy-preserving federated learning for genomic data

NEED:

- How do we conduct machine learning across data silos without sharing the source data?
- How do we know the machine learning models are private?
- What special considerations are there when conducting PPFL on genomic data?

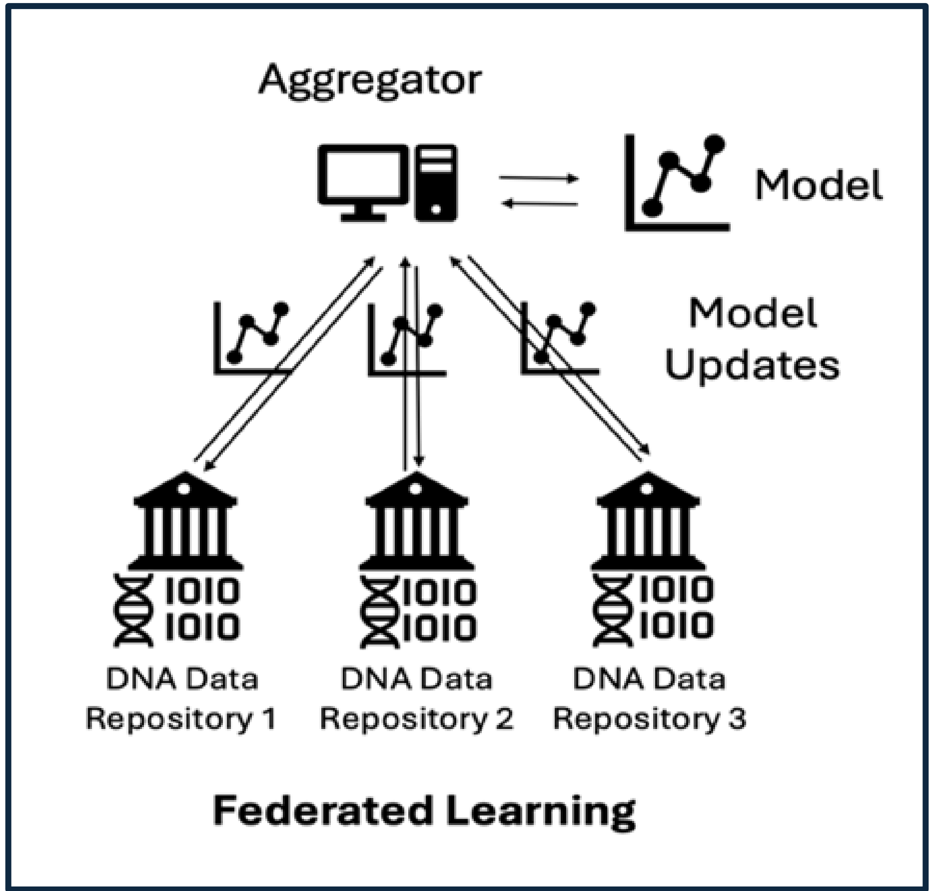
Search: NIST PPFL

GOALS:

- Create benchmark datasets
- Develop a demonstration PPFL network
- Red team privacy preserving techniques

OUTCOMES:

- Interesting *real,* datasets to benchmark algorithms
- Functioning PPFL cloud environment
- Red teaming exercise
- Research on what works and why
- *Coming soon:*
  - Publish code base and datasets
  - Accepting partners to run experiments



# Genomics PPFL Testbed

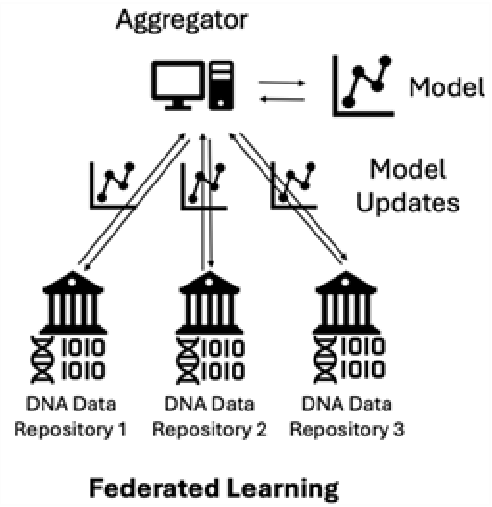


<u>Soybean traits</u>
20 features
50k SNPs
15k individuals

<u>Dog breed</u>
Hundres of dogs in different breeds and groups
10 features
50k SNPs

<u>Interesting data partitions</u>

<u>Easy experiment configuration</u>




<u>Easy to analyze evaluation resports</u>

- ML accuracy
- MIA

# Genomics PPFL Testbed

```
{ "model_type": "dpcnn",
  "num_partitions": 1,
  "partition_id": 0,
  "client_id": 0,
  "min_fit_clients": 1,
  "min_available_clients": 1,
  "min_evaluate_clients": 1,
  "n_models": 1,
  "num_rounds": 1,
  "num_cpus": 4,
  "seed": 673 ,
  "test_fraction": 0.3,
  "epochs": 100,
  "learning_rate": 0.003,
  "batch_divisor": 40,
  "weight_decay": 0.0001,
  "optimizer": "sgd",
  "opacus_secure_mode": false,
  "epsilon": 0.2,
  "delta": 1e-05,
  "max_grad_norm":1.0,
  "data_dir":"../data/gpd_scc",
  "data_partitions_file":"../data/ppfl_SCC_c4c5_5clients_2025_01_14.npz",
  "output_dir": "reports/gpd_scc_reports"
}
```

# Results from the Genomics Privacy-Preserving Federated Learning 2025 Red Teaming Event

**Christine Task**

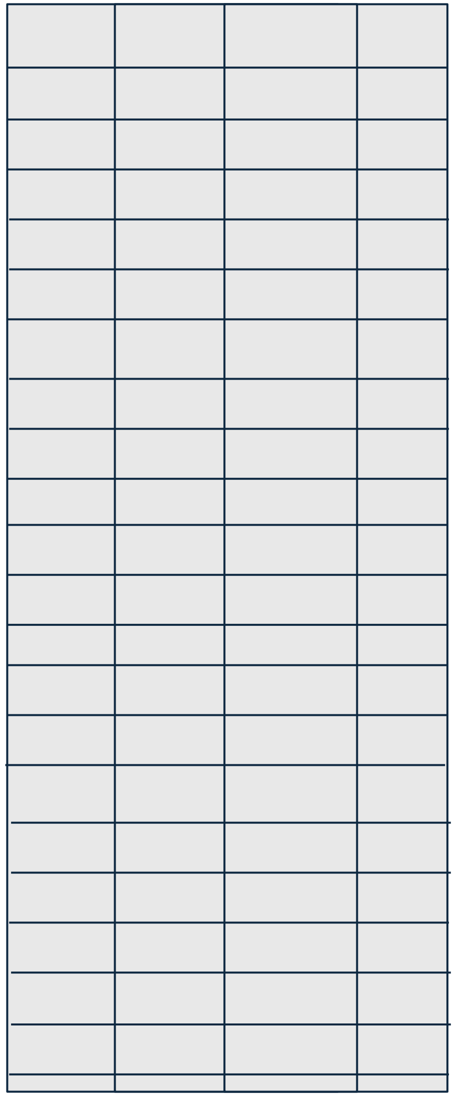
Director of Privacy Solutions and Synthetic Data

Knexus Research

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

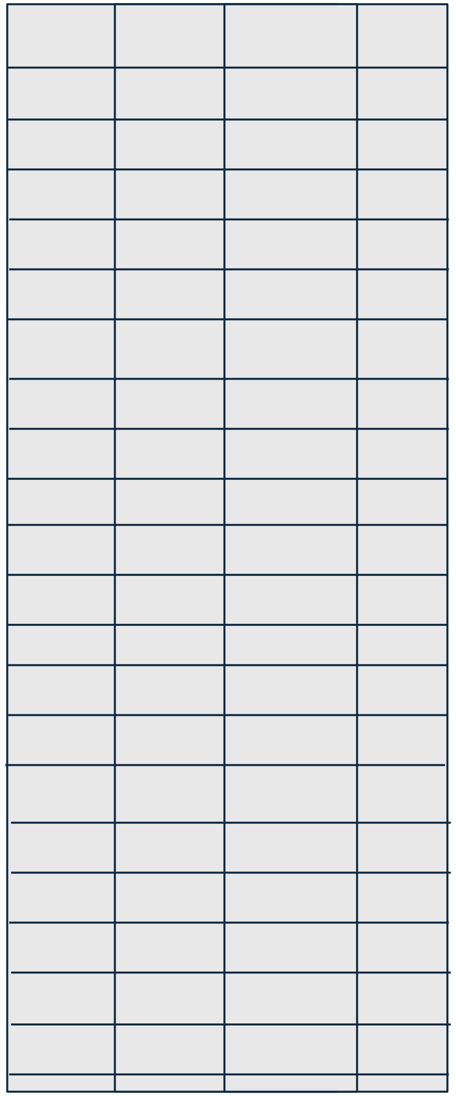
# What’s up with Genetic Data?

Differential privacy operates best with data that’s skinny (few dimensions) and tall (many individuals)



# What’s up with Genetic Data?

Differential privacy operates best with data that’s skinny (few dimensions) and tall (many individuals)



Genetic Data, especially for humans in a medical context, is *very* wide and short.

| Unnamed: 0 | Gm02_44000013 | Gm02_44000027 | Gm02_44000028 | Gm02_44000051 | Gm02_44000105 | Gm02_44000208 | Gm02_44000808 | Gm02_44000845 | Gm02_44000208 | Gm02_44000808 | Gm02_44000845 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| For | A/A | T/T | C/C | G/G | G/G | T/T | C/C | A/A | T/T | C/C | A/A |
| HN001 | A/A | T/T | C/C | G/G | G/G | T/T | C/C | A/A | T/T | C/C | A/A |
| HN002 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G | C/C | G/G | G/G |
| HN003 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G | C/C | G/G | G/G |
| HN006 | A/A | T/T | C/C | G/G | G/G | T/T | C/C | A/A | T/T | C/C | A/A |
| HN007 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G | C/C | G/G | G/G |
| HN008 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G | C/C | G/G | G/G |
| HN009 | A/A | T/T | C/C | G/G | G/G | T/T | C/C | A/A | T/T | C/C | A/A |
| HN011 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G | C/C | G/G | G/G |
| HN017B | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G | C/C | G/G | G/G |

There may be 50 or 100 individuals with a condition of interest in a given hospital system, and raw genetic data has **millions of columns**.

# What's up with Genetic Data?

Genetic Data, especially for humans in a medical context, has millions of columns and tens of rows. It's also not that interesting: It has low entropy. The vast majority of values will match the (publicly known) default value for that gene in the genome.

| Unnamed: 0 | Gm02_44000013 | Gm02_44000027 | Gm02_44000028 | Gm02_44000051 | Gm02_44000105 | Gm02_44000208 | Gm02_44000808 | Gm02_44000845 |
|---|---|---|---|---|---|---|---|---|
| For | A/A | T/T | C/C | G/G | G/G | T/T | C/C | A/A |
| HN001 | A/A | T/T | C/C | G/G | G/G | T/T | C/C | A/A |
| HN002 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G |
| HN003 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G |
| HN006 | A/A | T/T | C/C | G/G | G/G | T/T | C/C | A/A |
| HN007 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G |
| HN008 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G |
| HN009 | A/A | T/T | C/C | G/G | G/G | T/T | C/C | A/A |
| HN011 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G |
| HN017B | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G |
| HN018 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G |
| HN021 | A/A | T/T | C/C | G/G | G/G | T/T | C/C | A/A |
| HN023 | C/C | C/C | A/A | A/A | C/C | C/C | C/G | A/G |

# What’s up with Genetic Data?

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Genetic Data, especially for humans in a medical context, has millions of columns and tens of rows.

It’s also not that interesting: It has low entropy. The vast majority of values will match the (publicly known) default value for that gene in the genome.

**Q1:** What’s a meaningful reconstruction attack against **something everyone already mostly knows?**

| Unnamed: 0 | Gm02_44000013 | Gm02_44000027 | Gm02_44000028 | Gm02_44000051 | Gm02_44000105 | Gm02_44000208 | Gm02_44000808 | Gm02_44000845 |
|---|---|---|---|---|---|---|---|---|
| For | A/A | T/T | C/C | G/G | G/G | T/T | C/C | A/A |
| HN001 | A/A | T/T | C/C | G/G | G/G | T/T | C/C | A/A |
| HN002 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G |
| HN003 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G |
| HN006 | A/A | T/T | C/C | G/G | G/G | T/T | C/C | A/A |
| HN007 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G |
| HN008 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G |
| HN009 | A/A | T/T | C/C | G/G | G/G | T/T | C/C | A/A |
| HN011 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G |
| HN017B | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G |
| HN018 | C/C | C/C | A/A | A/A | C/C | C/C | G/G | G/G |
| HN021 | A/A | T/T | C/C | G/G | G/G | T/T | C/C | A/A |
| HN023 | C/C | C/C | A/A | A/A | C/C | C/C | C/G | A/G |

# PPFL Red Teaming

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

**March – June: Phase 1 Soybean Data**

**Basic Problem (Seed Coat Columns): Membership Inference**

- March 10 Problem Launch
- April 11st First Leaderboard Deadline
- May 1st Second Leaderboard Deadline
- June 2nd Third Leaderboard Deadline
- July 1st Final Leaderboard Deadline

**Advanced Problem: Reconstruction**

- May 2nd Problem Launch
- June 2nd First Leaderboard Deadline
- July 1st Final Leaderboard Deadline

**June – July: Phase 2 Dog Data**

**Basic Problem: Membership**

- June 3 Problem Launch
- July 15th Final Leaderboard Deadline

**Advanced Problem: Reconstruction**

- June 3 Problem Launch
- July 15th Final Leaderboard Deadline

# Individually Identifiying Genes

The vast majority of values will match the (publicly known) default value for that gene in the genome.

**Q1:** How do we run a meaningful reconstruction attack against something everyone already knows?

**Individually Identifying Genes for Evaluating Meaningful Reconstruction Attacks**

1. **High Variance:** Only consider columns with high entropy; genes where many individuals in the population have variations. This reduces the millions of columns in the original data to a few hundred.

2. **Mutually Independent Clusters:** Look at the pairwise conditional dependence between genes and use a graph partitioning algorithm to cluster the genes into associated subsets that are relatively independent of each other.

3. **Identifying Genes**: Take the columns from the highest entropy, most independent clusters. These ~30 genes are sufficient to uniquely identify more than 80% of the individuals in the data. ***We award extra points if you're able to reconstruct these.***

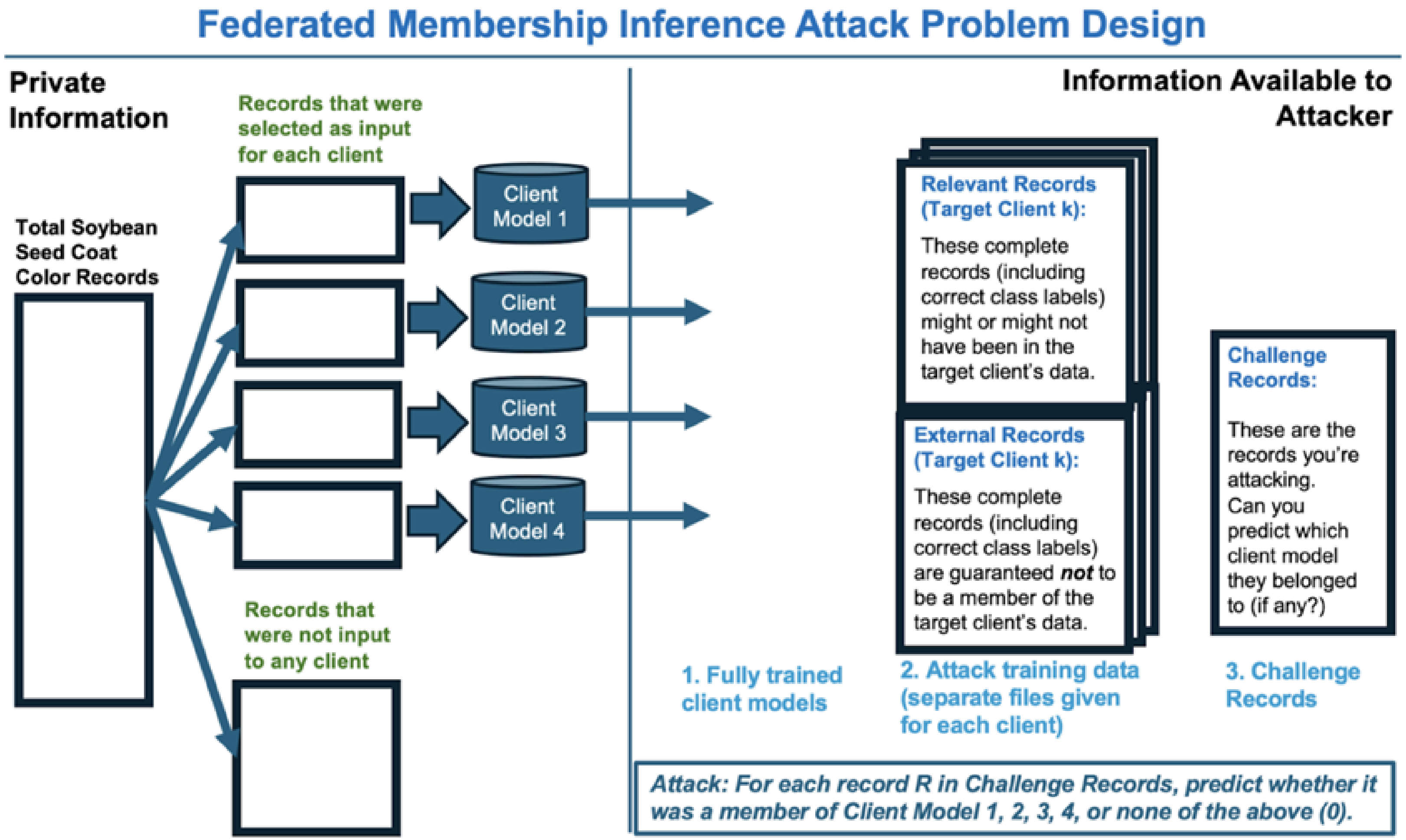
# Other Attacks

The vast majority of values will match the (publicly known) default value for that gene in the genome.

**Q1:** How do we run a meaningful reconstruction attack against something everyone already knows?

**<u>Other Meaningful Attacks:</u>**

**Membership Inference**



# Other Attacks

The vast majority of values will match the (publicly known) default value for that gene in the genome.

**Q1:** How do we run a meaningful reconstruction attack against something everyone already knows?

**<u>Other Meaningful Attacks:</u>**

**Membership Inference**

**Subpopulation Inference (“Hound Hunt”)**

**Local Client Training**: When clients are performing local training, they repeatedly select random batches of records from their local data and use them to compute update gradients. Batch sizes are either 5 or 40 records for our models.

# Preparing the Models to be Attacked

**Q2:** How do we train task models with differential privacy on very wide, short data?

**Training the PPFL Classification Task Models with Current Medical FL Practices:**

# Preparing the Models to be Attacked

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

**Q2:** How do we train task models with differential privacy on very wide, short data?

**Training the PPFL Classification Task Models with Current Medical FL Practices:**

1. **Reduced Genetic Data:** Reduced data has thousands of columns rather than millions. Mix of uniform random subselection and data analysis to identifying columns relevant to the task.

2. **One Hot Enoding:** Expand the number of columns back up by a factor of 4 using one-hot encoding to transform categorically data (A/G) into a set of binary columns (0100) with deterministic dependencies.

# Preparing the Models to be Attacked

**Q2:** How do we train task models with differential privacy on very wide, short data?

**<u>Training the PPFL Classification Task Models with Current Medical FL Practices:</u>**

1. **Reduced Genetic Data:** Reduced data has thousands of columns rather than millions. Mix of uniform random subselection and data analysis to identifying columns relevant to the task.

2. **One Hot Enoding:** Expand the number of columns back up by a factor of 4 using one-hot encoding to transform categorically data (A/G) into a set of binary columns (0100) with deterministic dependencies.

3. **Create Realistic Client Subpopulations:** To create realistically heterogeneous client data sets, we clustered data along the 5th and 6th PCA axes to create a subtly biased (non-iid) distribution between clients. Some clients data was very small (~50), and others larger (~200)

**4. Require minimum task accuracy:** We set epsilon so the accuracy on the training data was at least 70% for all but the smallest client. Even with overfitting, this meant epsilon ≧ 5 on the simplest tasks and data (soybean data), and epsilon ≧ 20 on the more complex tasks (dog data)

# Observations During Design

**Q3:** What's happening with federated learning on genomics data in current practice?

**Observations from Preparing the Challenge Models:**

# Observations During Design

**Q3:** What's happening with federated learning on genomics data in current practice?

**Observations from Preparing the Challenge Models:**

1. **We weren't getting exceptional performance on task models even before we added noise for privacy.** Federated learning on heterogeneous, wide, short, low entropy, one-hot encoded data is hard. Subsequent conversations haven't turned up anyone who's actually used federated learning on genomics data to learn new information about a medical condition that impacted clinical recommendations or practice.

# Observations During Design

**Q3:** What's happening with federated learning on genomics data in current practice?

**Observations from Preparing the Challenge Models:**

1. **We weren't getting exceptional performance on task models even before we added noise for privacy.** Federated learning on heterogeneous, wide, short, low entropy, one-hot encoded data is hard. Subsequent conversations haven't turned up anyone who's actually used federated learning on genomics data to learn new information about a medical condition that impacted clinical recommendations or practice.

2. **As epsilon decreased, we lost utility before we gained privacy:** Long after added noise caused models to completely fail their classification task, it was still possible to perform successful membership inference attacks against them. ***They didn't have utility or privacy***.

# Observations During Design

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

**Q3:** What's happening with federated learning on genomics data in current practice?

**Observations from Preparing the Challenge Models:**

1. **We weren't getting exceptional performance on task models even before we added noise for privacy.** Federated learning on heterogeneous, wide, short, low entropy, one-hot encoded data is hard. Subsequent conversations haven't turned up anyone who's actually used federated learning on genomics data to learn new information about a medical condition that impacted clinical recommendations or practice.

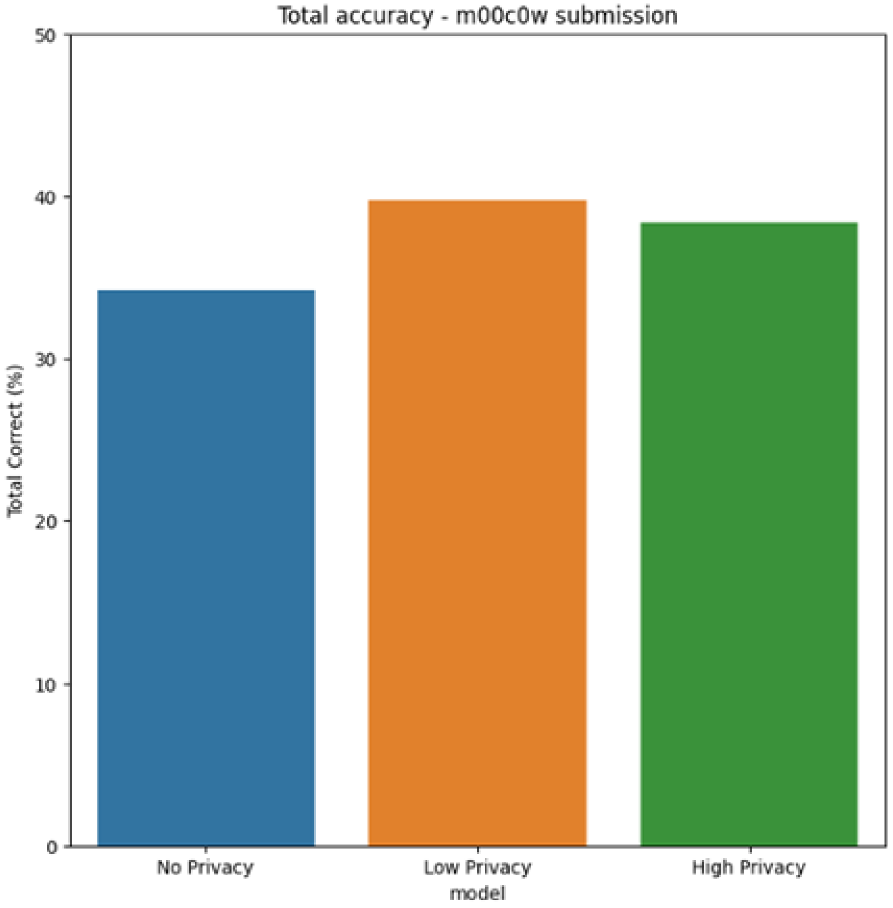
2. **As epsilon decreased, we lost utility before we gained privacy:** Long after added noise caused models to completely fail their classification task, it was still possible to perform successful membership inference attacks against them. ***They didn't have utility or privacy***.

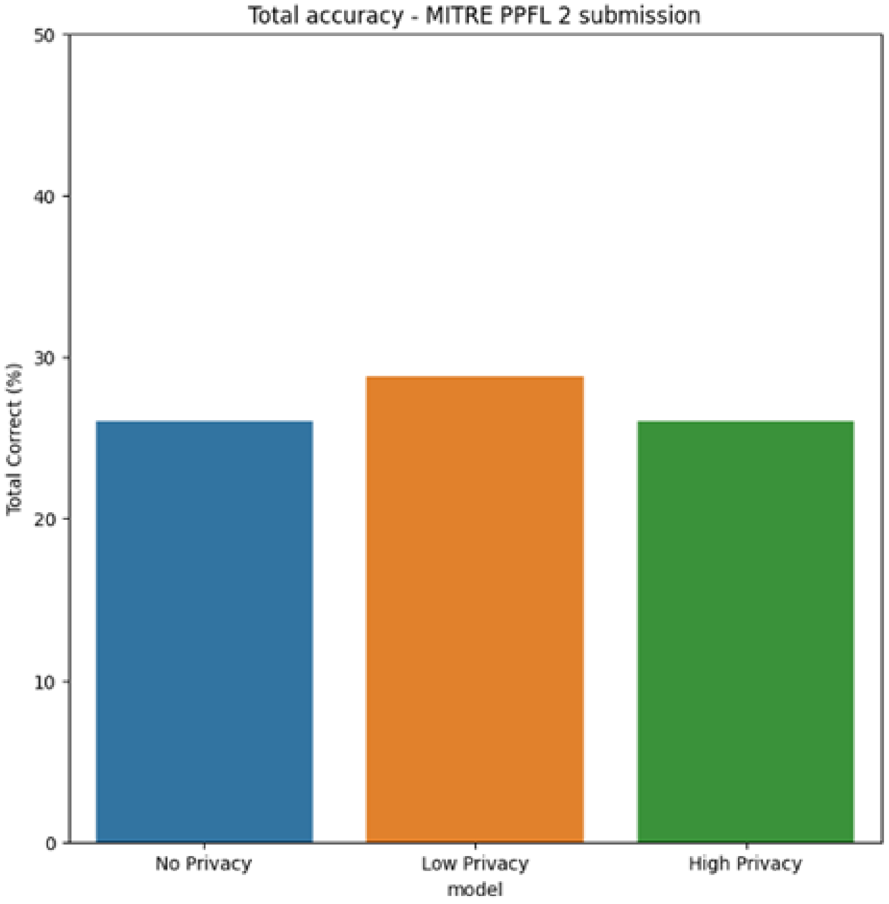
3. **Meaningful reconstruction is hard:** Well, they still had some privacy. Regardless of the amount of noise added (or none at all), reconstructing individually identifying genes or inferring whether a gradient included a member of a targeted subpopulation usually saw low success.

# What Happened?

Epsilon 10, 200, and inf had very different levels of utility on the task performance, but all produced similar levels of resistance against the most successful membership attacks.

The most successful attack against the most vulnerable client model achieved an 80% success rate.

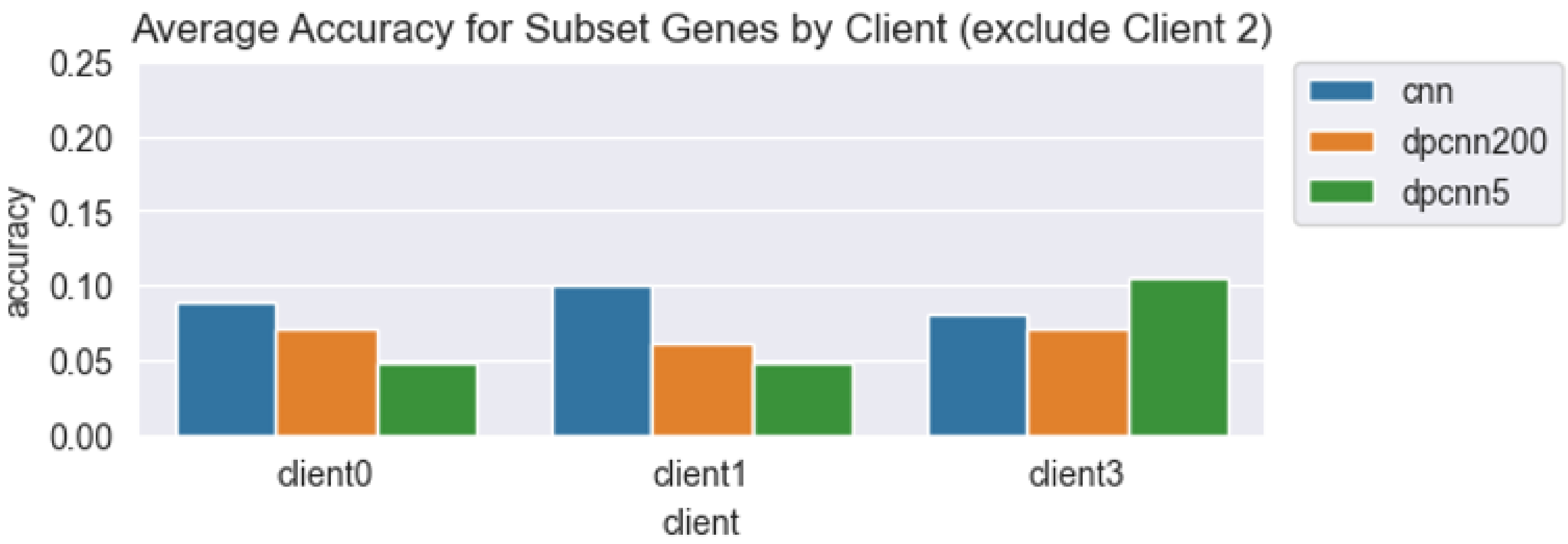


# What Happened?

Epsilon 5, 200, and inf produced somewhat different levels of resistance against reconstruction of individually identifying genes.

But that accuracy was very low, **< 10%**.

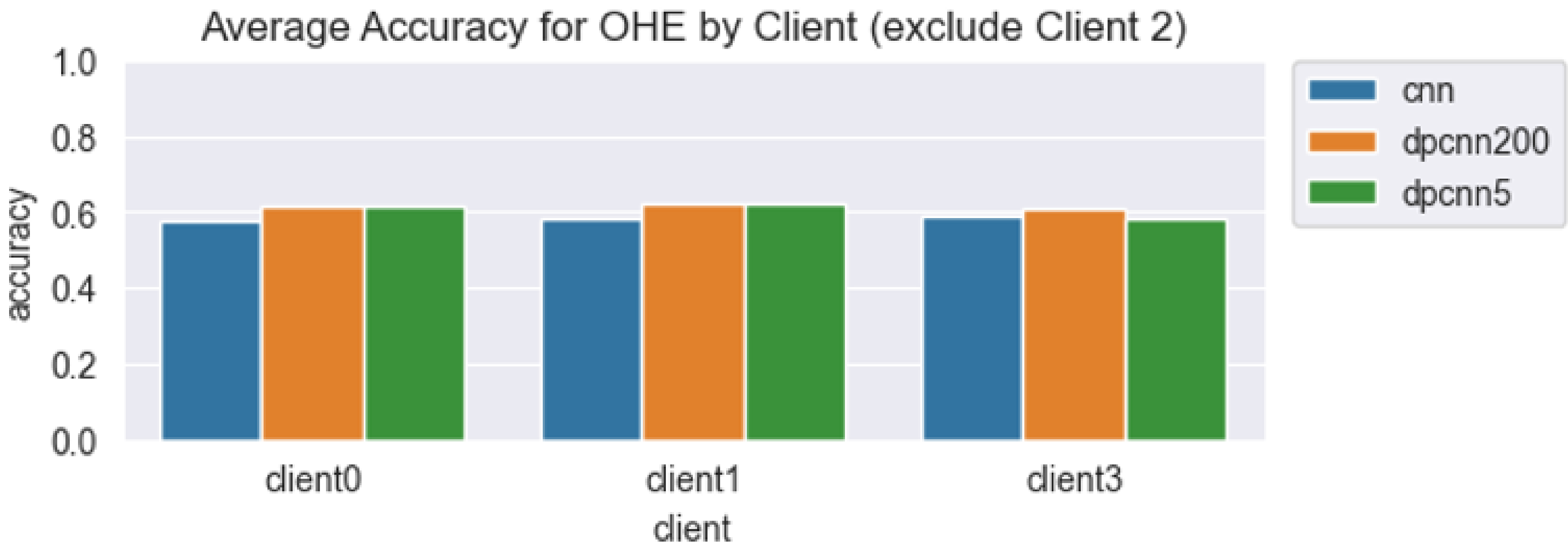


# What Happened?

Epsilon 5, 200, and inf produced somewhat different levels of resistance against reconstruction of individually identifying genes.

But that accuracy was very low, < 10%.

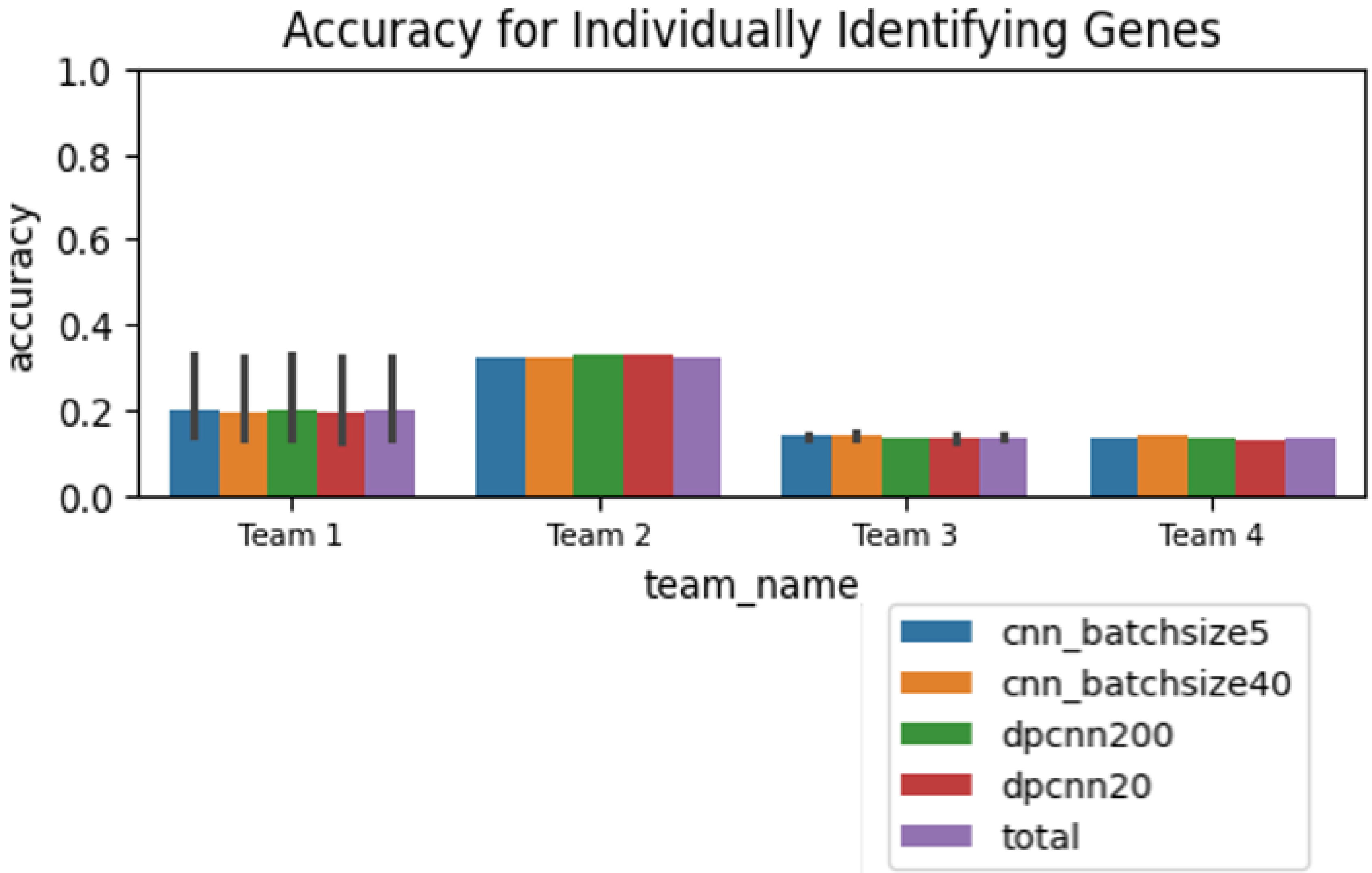
If we gave them credit for getting all the zeros right in the one hot encoded data, then accuracy neared the percentage of zeros.



# What Happened?

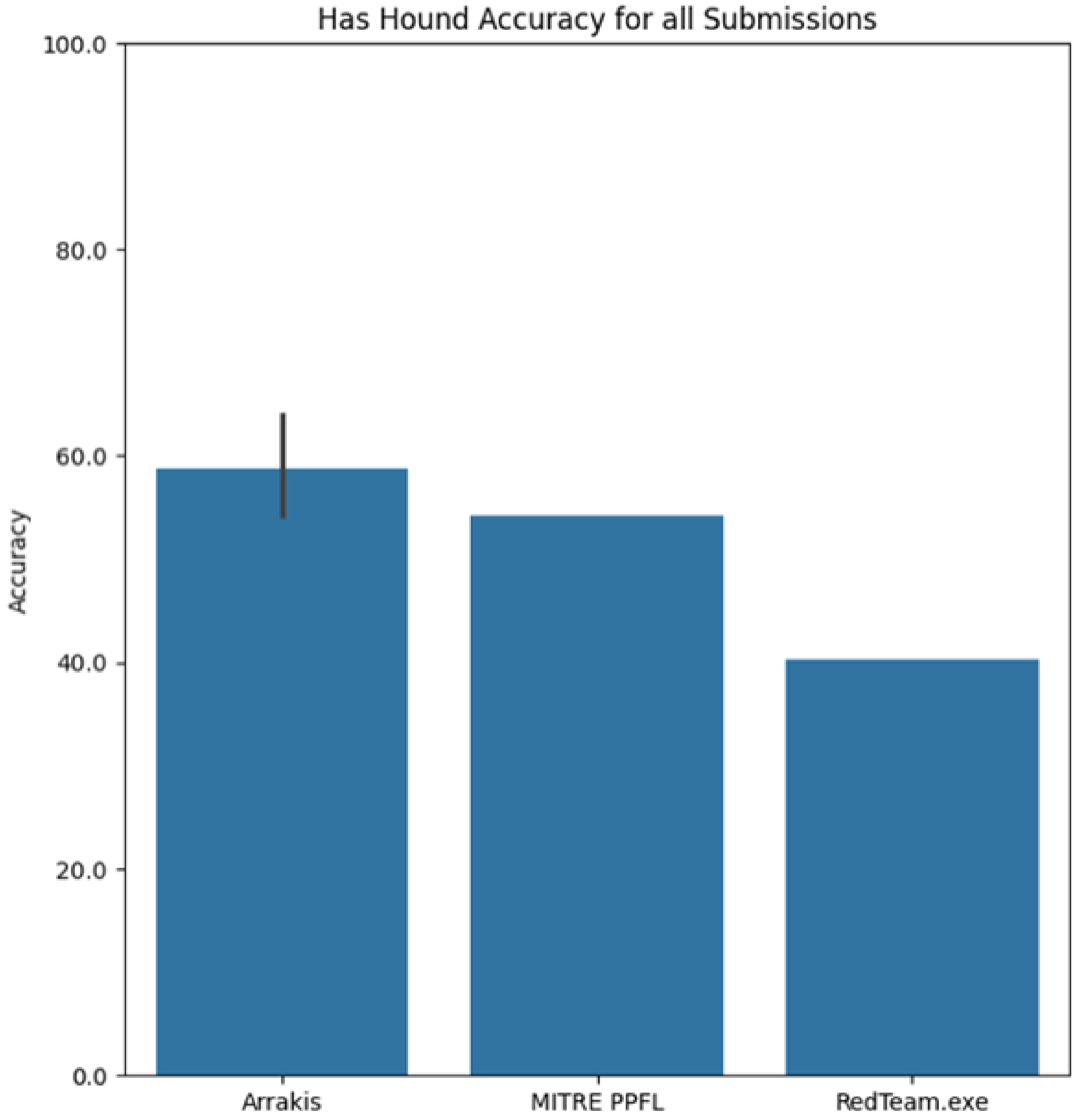
On our second reconstruction problem (with dog data), we saw epsilon 20, 200, and inf all produce the same levels of resistance against reconstruction of individually identifying genes.

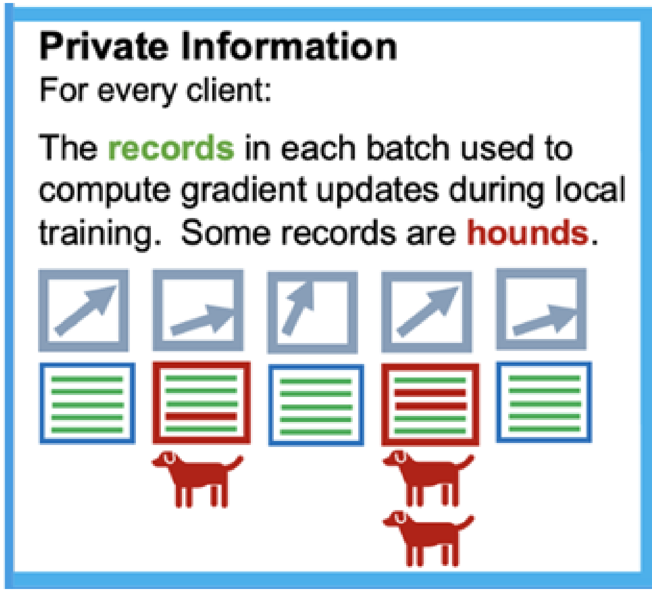
But this time we handed it (separately) to Claude and Mitre, and they each addressed the data representation problem (maintaining OHE constraints and reducing dimensionality) getting up to **40% accuracy.**



# What Happened?

Guessing whether or not a gradient included a hound (target subpopulation) was still basically a 50/50 shot though.



# Conclusions and Next Steps

**Q4:** What can we take away from this? *Differentially privacy didn't seem to do much at epsilons that supported minimum model task accuracy.*

The same lesson from the rest of the conference: **We need the data to be represented in a space that helps us maintain the signal as we add noise.**

A better representation would likely also improve the overall feasibility of federated learning on human genomics data. Epsilons can be smaller (and protection can be stronger) if the learning task itself isn't so difficult that any added noise immediately eliminates already tenuous model utility.

So rather than one hot encoding, let's find clever ways to significantly reduce the dimensionality of genomics data before attempting differentially private federated learning over it.

Remember that default genomes are public data, as is a lot of other information about genetic analysis.

# Discussion Topics

**1. Planned Data Uses**

What are some specific examples of analyses, AI applications, reports or other tasks for the data that you would use PETs with?

**2. Stakeholder Engagement and Decision-Making**

What are the privacy considerations for the data? Confidential columns/variables, data use agreement or legal restrictions, etc?

**3. Future Aspirations**

If you had a magic wand, what would you do in terms of collaboration and data sharing you're unable to accomplish now?

**4. Cross-Sector Collaboration**

What other sectors may be interested in interacting with the PETs Testbed?

# Dioptra

## An Open-Source Test Platform for AI

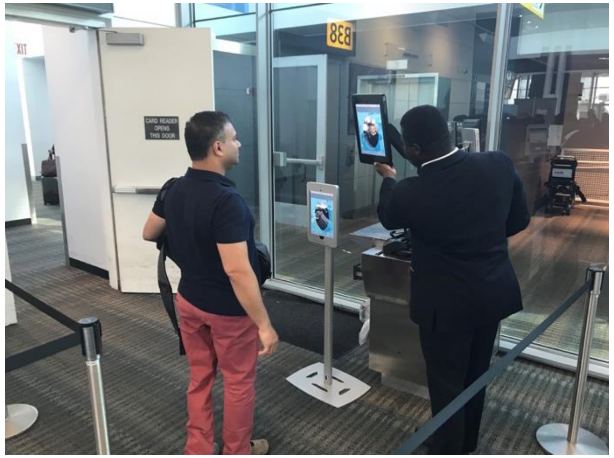
**Harold Booth**
Computer Scientist, NIST
**Keith Manville**
Principal AI Security Engineer, MITRE/NCCoE
**Dr. James Glasbrenner**
Principal AI Engineer, MITRE/NCCoE

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

# Test AI Systems in Context

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

**Security Checkpoint**

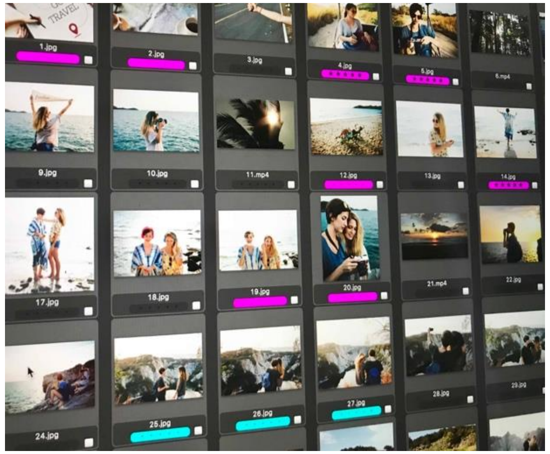


Controlled environment

**Automated Driving**

Open environment

**Image Forensics**



Open environment

- What is the task?
- Where (and when) can you control the environment?
- What are potential adversarial goals?
- What components does the AI system depend on?
- What functions depend on the AI system?

# Dioptra as a Test Platform for Assessing Characteristics of AI Systems

## Dioptra is NIST's open-source software test platform for assessing characteristics of artificial intelligence

From *Cybersecurity Framework Profile for Artificial Intelligence (Cyber AI Profile)*: https://csrc.nist.gov/pubs/ir/8596/iprd

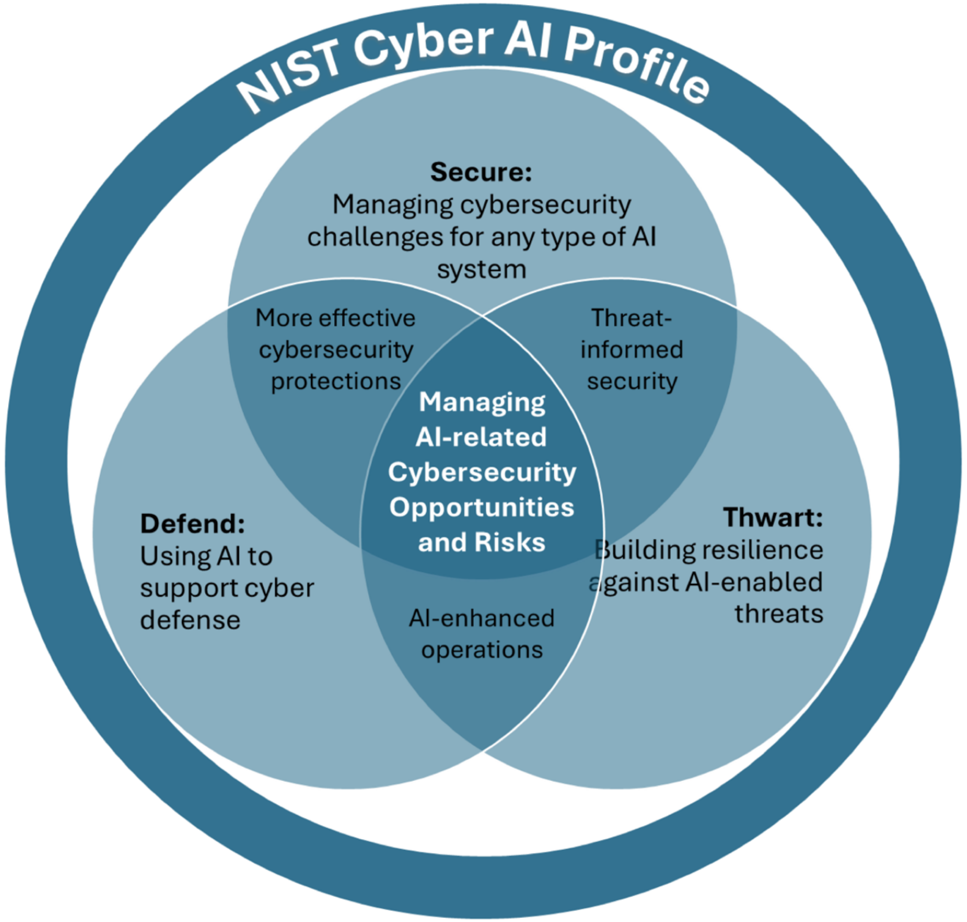



**Fig. 11: Relationship Between Cyber AI Profile Focus Areas**

# Dioptra as a Test Platform for Assessing Characteristics of AI Systems

## Dioptra is NIST's open-source software test platform for assessing characteristics of artificial intelligence

From *Cybersecurity Framework Profile for Artificial Intelligence (Cyber AI Profile)*:
https://csrc.nist.gov/pubs/ir/8596/iprd

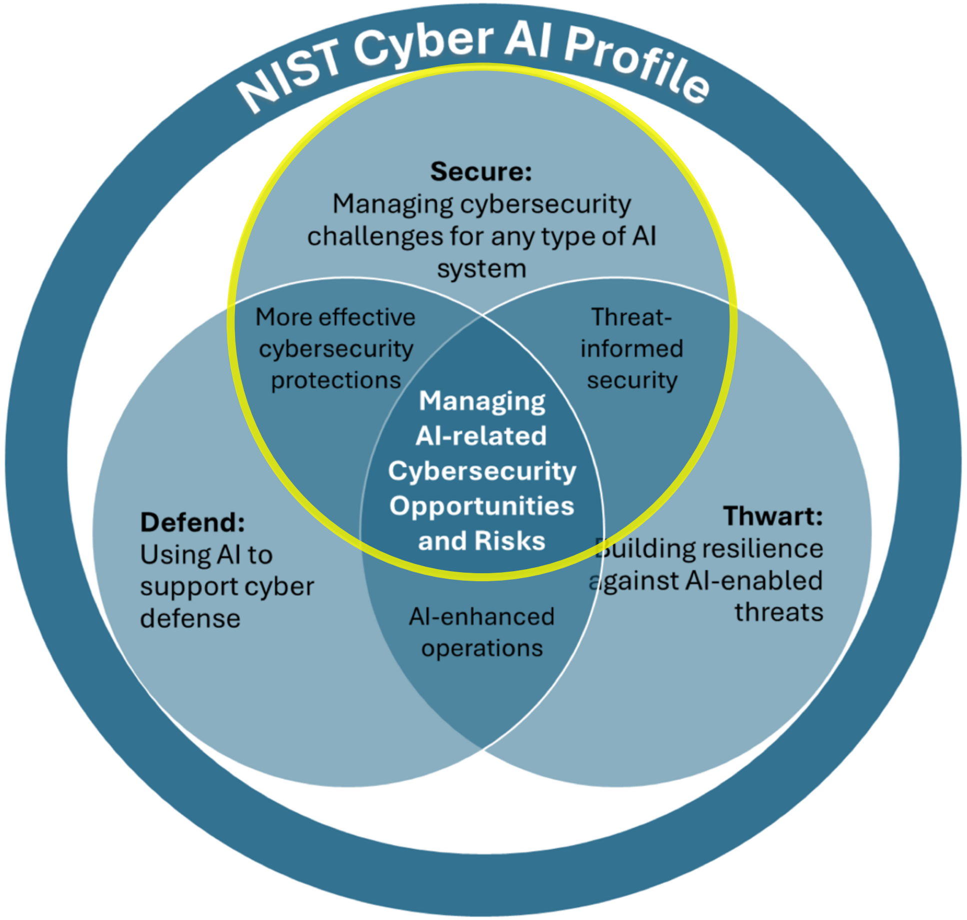
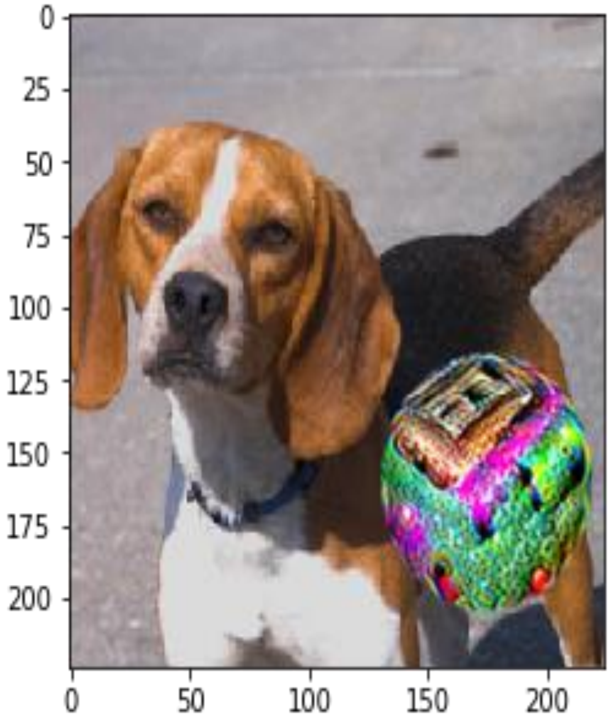



Fig. 11: Relationship Between Cyber AI Profile Focus Areas

# Motivating Example: Evasion Attacks

## Patch Evasion

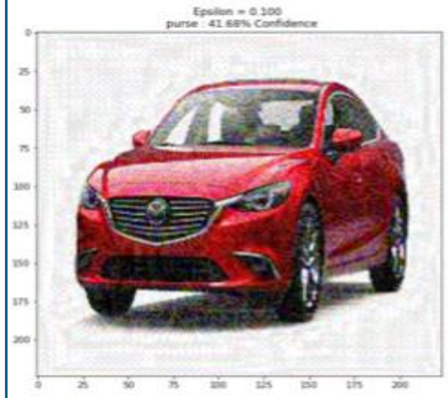



**Predictions:**

| | |
|---|---|
| Toaster | 30% |
| Beagle | 11% |
| Spatula | 8% |

## Perturbation Evasion


sports car: 75%

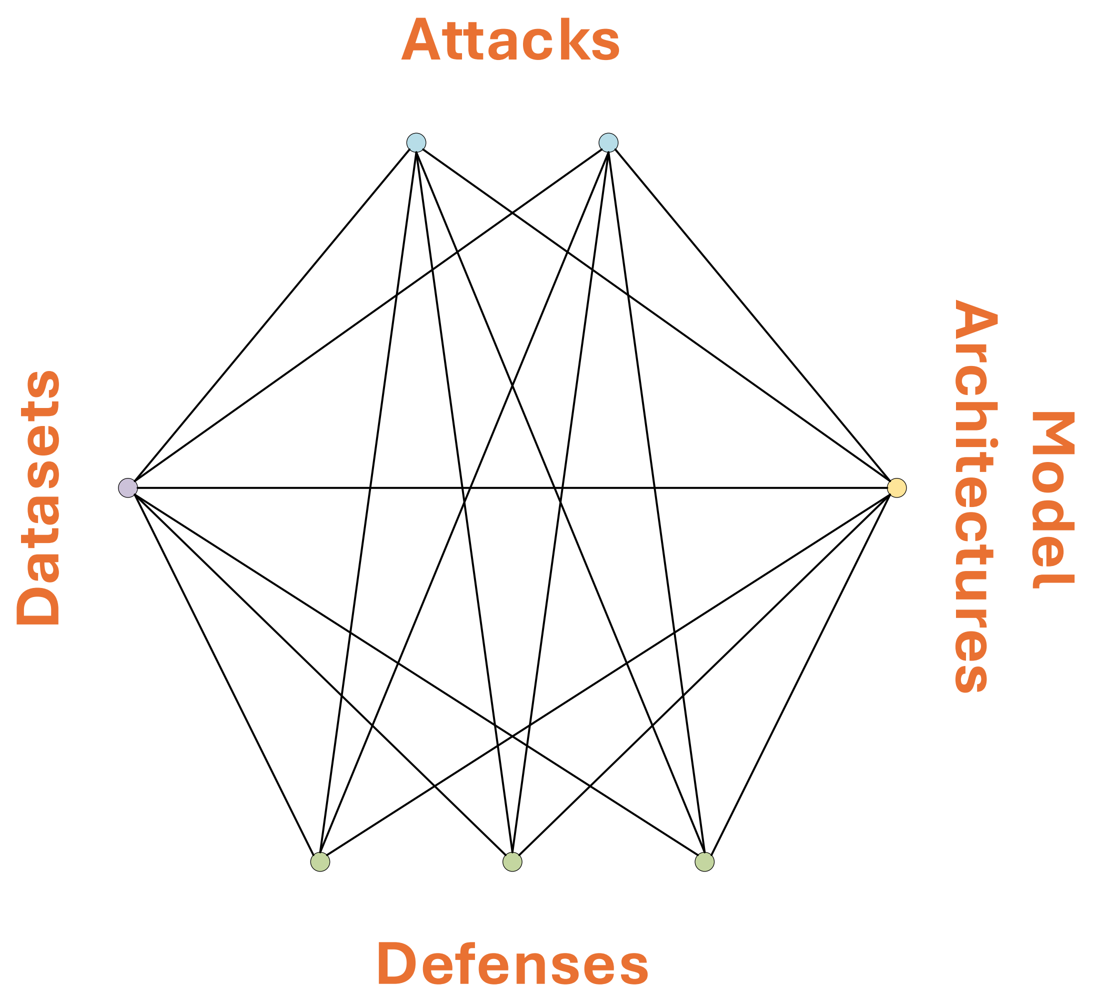



purse: 41%

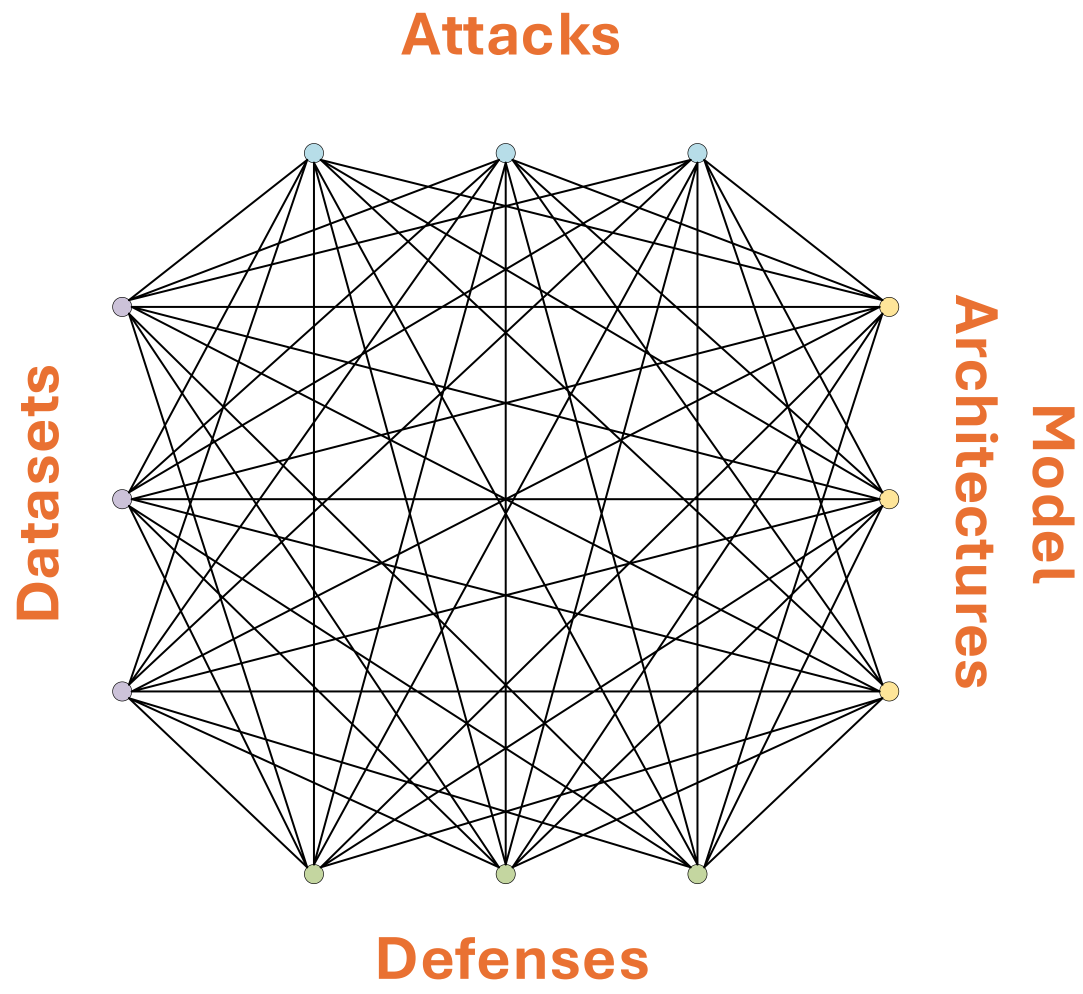


Images From ImageNet and Fruits360 Dataset

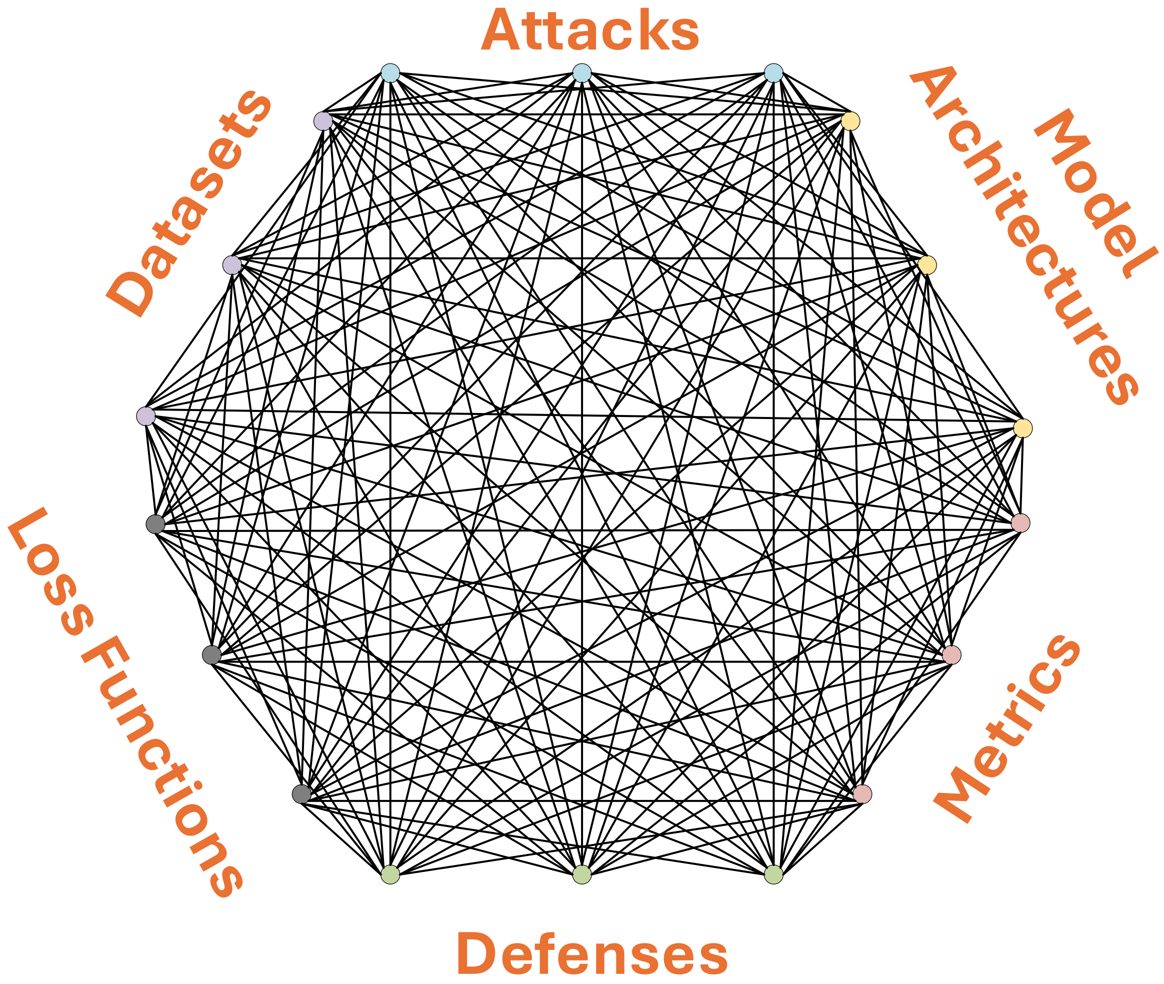
# The Complexity of Problem Exploration



# The Complexity of Problem Exploration



# The Complexity of Problem Exploration



# Key Properties in an AI Test Platform

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

**Reproducible:**

Dioptra automatically creates snapshots of resources so experiments can be reproduced and validated

**Traceable:**

The full history of experiments and their inputs/outputs are tracked, enabling generation of a detailed bill of materials

**Extensible:**

Support for expanding functionality and importing existing Python packages via a plugin system

**Interoperable:**

A type system promotes interoperability between plugins

**Modular:**

New experiments can be composed from modular components in a simple configuration file

**Secure:**

Dioptra provides a multi-user environment with authentication

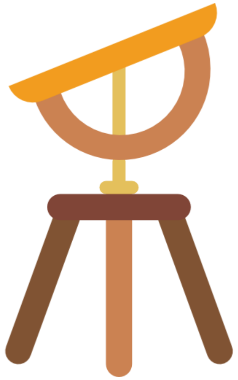
**Interactive:**

Users can interact with Dioptra via an intuitive web interface or Python client

**Shareable and Reusable:**

Dioptra components can be shared and re-used across deployments

# Flexible and Modular Evaluation



**Dioptra**

Image: Flaticon.com/Smashicons

- Shallow Net
- AlexNet
- LeNet
- ResNet50
- VGG16
- …

**Model Architecture**

- Patch augmentation
- Poison Frogs
- Adversarial training
- …

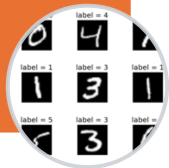
**Data Augmentation**

- Spatial smoothing
- Defensive distillation
- …

**Inference pre-processing**

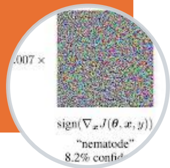
- MNIST
- Fruits360
- ImageNet
- …

**Dataset**



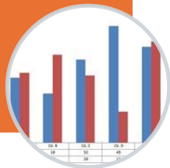
- Fast Gradient Method
- Pixel Threshold
- Patch
- Membership Inference
- …

**Attack on trained model**



- Clean accuracy
- Adversarial accuracy
- Robustness radius
- …

**Metric**



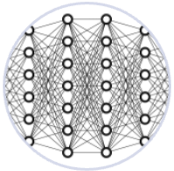
# Formative Use Cases for Dioptra

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

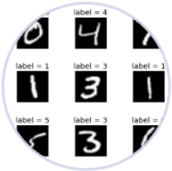
- Researchers
- 2nd Party Testing
- 1st Party Testing
- 3rd Party Testing
- Evaluations / Challenges

# What Do Users Need to Bring to Dioptra?





Models

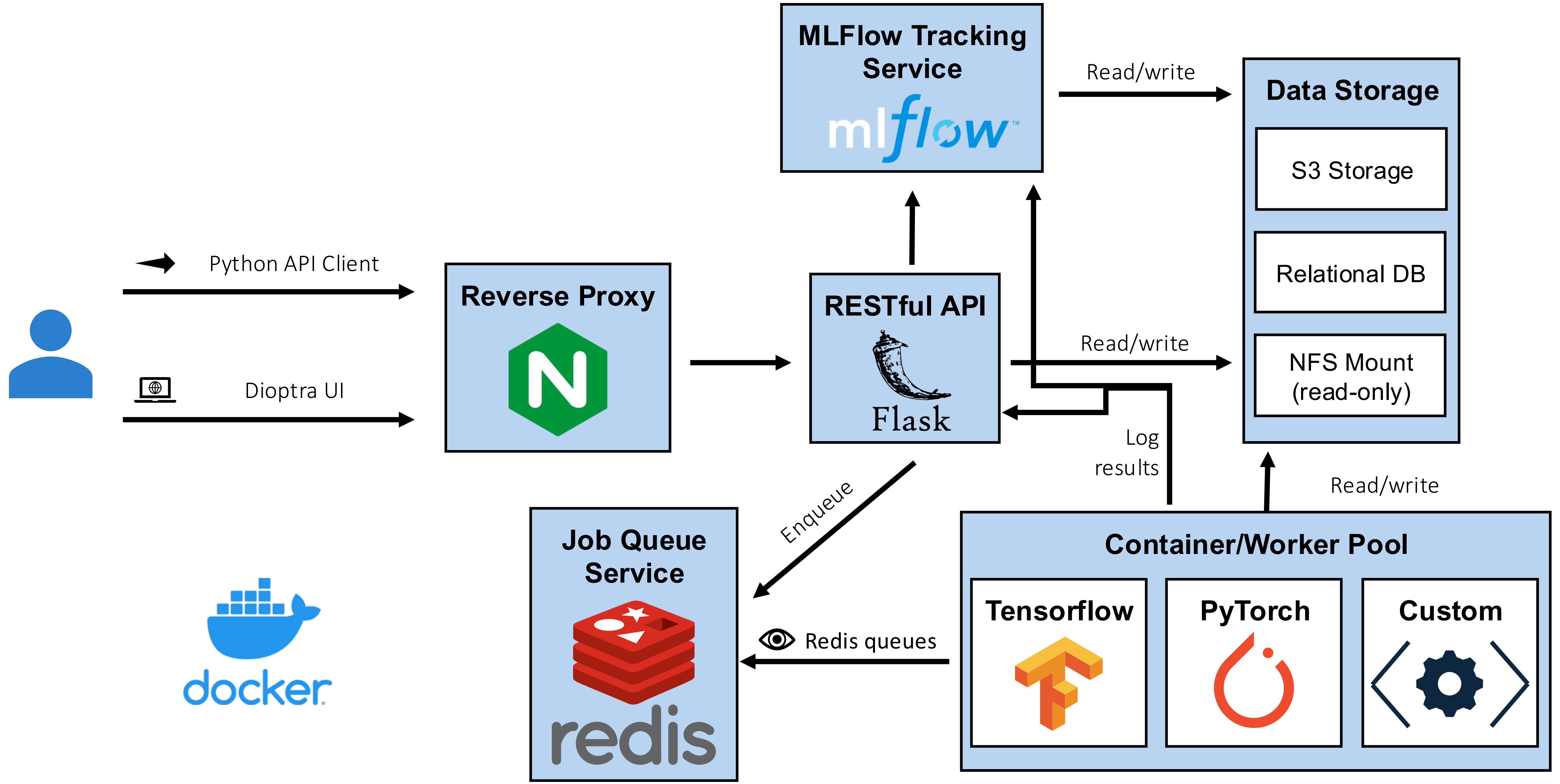


Data

Compute

Code

# Dioptra Architecture



# Demo

- Evasion Attacks

# Genomic Data Analysis

- Current work focuses on using Dioptra to train and test Stratomod - a NIST genomic variant calling error classification model (https://www.nature.com/articles/s42003-024-06981-1)

- In-progress work also includes a demonstration of tracking provenance for a genomic data analysis pipeline

# Repository:

https://github.com/usnistgov/dioptra

Questions: dioptra@nist.gov

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

# Webinar Take Aways

- NIST is addressing genomic data cybersecurity and privacy risks through:
  - Threat modeling -- see draft publications and engagement recordings at https://www.nccoe.nist.gov/projects/cybersecurity-and-privacy-genomic-data
  - PETs testbed -- https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/pets-testbed
  - AI evaluation infrastructure (Dioptra) -- https://pages.nist.gov/dioptra/

- You can help collaborate with us:
  - Submit comments to draft publications
  - Upon release, use PETs testbed to evaluate privacy metrics as well as test using Dioptra

- Be on the lookout for announcements of genomic use case demonstrations for Dioptra and the PETs testbed

# Webinar Closeout

Ron Pulivarti, NIST NCCoE

Justin Wagner, NIST MML

1. Join our Community of Interest (COI)
2. Stream Workshops & Webinars
3. Review Publications & other Resources

https://www.nccoe.nist.gov/genomics-cybersecurity

# THANK YOU

https://www.nccoe.nist.gov/projects/cybersecurity-and-privacy-genomic-data

genomic_cybersecurity_nccoe@nist.gov

nccoe.nist.gov

@NISTcyber