SPIA 586a: Machine Learning for Policy Analysis
Week 10: Privacy

Brandon Stewart[1]

Princeton

March 31, 2025

[1]Many thanks to Ian Lundberg and Matt Salganik.

Where We've Been and Where We're Going...

Where We've Been and Where We're Going...

- Last Week
  - found data

Where We've Been and Where We're Going...

- Last Week
  - found data
- This Week
  - privacy

Where We've Been and Where We're Going...

- Last Week
  - found data
- This Week
  - privacy
- Next Week
  - fairness and racial justice

Where We've Been and Where We're Going...

- Last Week
  - found data
- This Week
  - privacy
- Next Week
  - fairness and racial justice
- Long Run
  - target tasks → data sources → guiding values

Three Broad Categories of Values

Three Broad Categories of Values

- Privacy
  what information is exposed?

Three Broad Categories of Values

- Privacy
  what information is exposed?
- Fairness
  are gains and costs distributed equitably?

Three Broad Categories of Values

- Privacy
  what information is exposed?
- Fairness
  are gains and costs distributed equitably?
- Interpretability
  do we know what the machine algorithm is doing?

Core Challenges for Privacy

Core Challenges for Privacy

1) Anonymization is really, really hard.

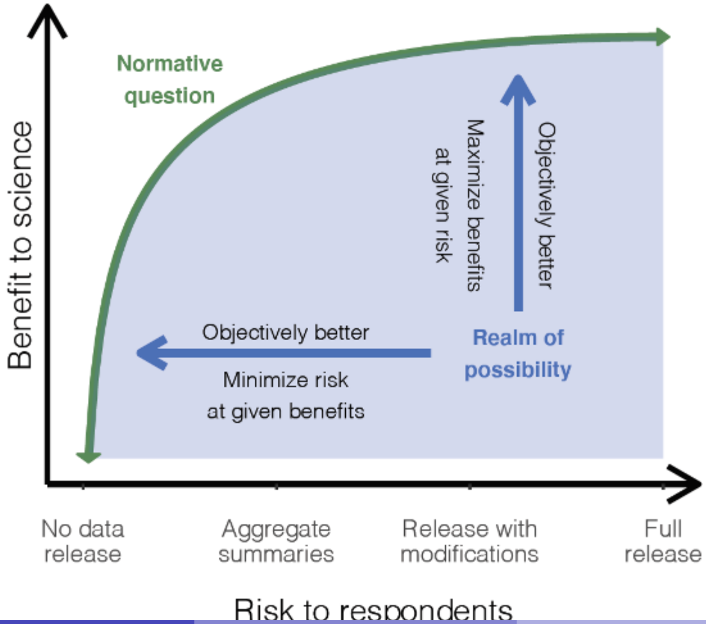
Core Challenges for Privacy

1) Anonymization is really, really hard.
2) It is hard to control access.

Core Challenges for Privacy

1) Anonymization is really, really hard.
2) It is hard to control access.
3) The public/private distinction isn't as clean as it once was.

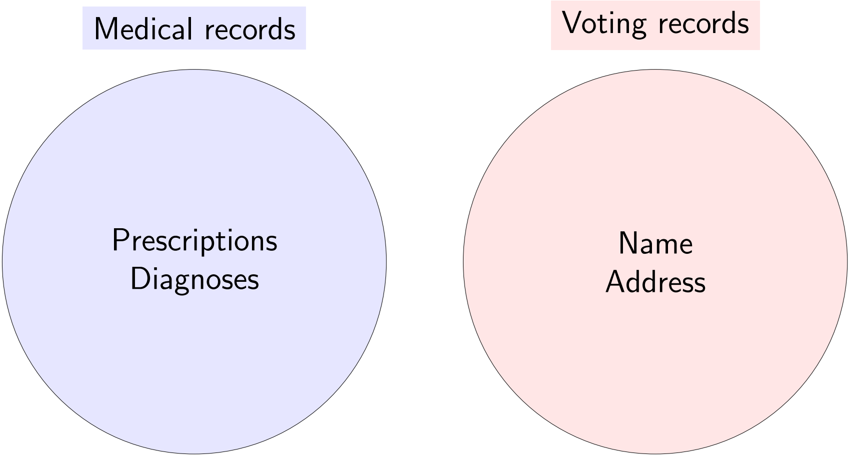
Tradeoffs



Benefit to science
Normative question
Maximize benefits at given risk
Objectively better
Objectively better
Minimize risk at given benefits
Realm of possibility
No data release
Aggregate summaries
Release with modifications
Full release
Risk to respondents

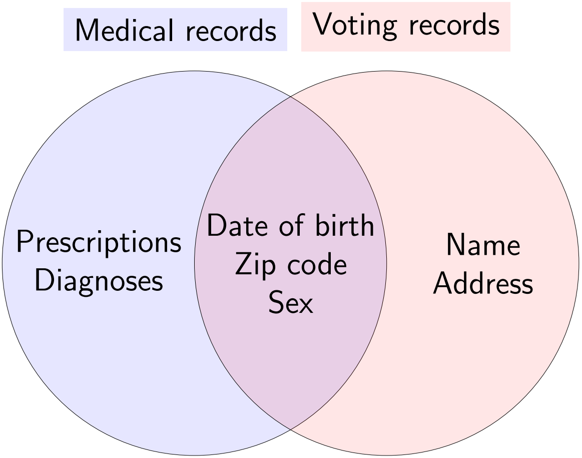
Challenge 1: Anonymization is Hard

Sweeney (1997) re-identified Massachusetts medical records



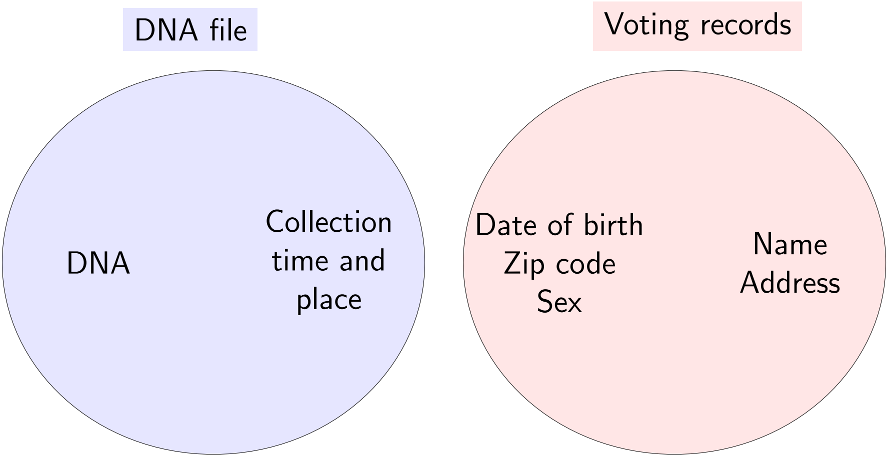
Slide Credit: Ian Lundberg and Matt Salganik

Sweeney (1997) re-identified Massachusetts medical records



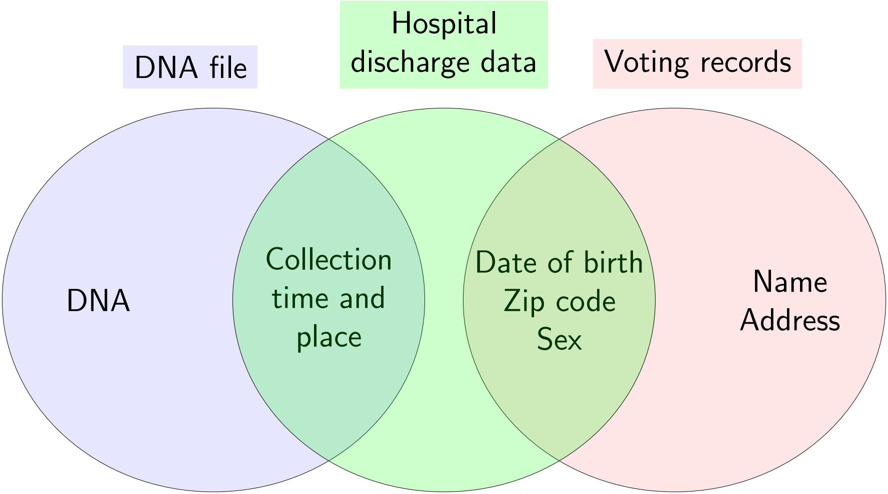
Slide Credit: Ian Lundberg and Matt Salganik

Malin and Sweeney (2004) re-identified genomics data



Slide Credit: Ian Lundberg and Matt Salganik

Malin and Sweeney (2004) re-identified genomics data



Slide Credit: Ian Lundberg and Matt Salganik

Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan and Vitaly Shmatikov
The University of Texas at Austin

dx.doi.org/10.1109/SP.2008.33

RYAN SINGEL SECURITY 12.17.09 04:29 PM

NETFLIX SPILLED YOUR BROKEBACK MOUNTAIN SECRET, LAWSUIT CLAIMS

"[M]ovie and rating data contains information of a more highly personal and sensitive nature [sic]. The member's movie data exposes a Netflix member's personal interest and/or struggles with various highly personal issues, including sexuality, mental illness, recovery from alcoholism, and victimization from incest, physical abuse, domestic violence, adultery, and rape." (Singel, 2009)

Differential Privacy

In some contexts, can make strong guarantees of privacy by releasing data with added noise.

Social Science One

SOCIAL SCIENCE ONE

Hosted by Harvard's Institute for Quantitative Social Science

BLOG CONTACT U

Building Industry-Academic Partnerships

Academic social scientists have more data than ever before to study human society, but a smaller proportion of existing data than at any time in history because most of it is now tied up inside companies.

Through information sharing, demonstration projects, collaborations, and other activities, we seek to reduce barriers to industry collaboration, unlock commercial information for public good in privacy protective ways, and enable both academics and companies to advance their separate and joint goals.

A Consortium of Research Centers

With leading social science research centers around the world,

Company Partnerships

We develop models for industry collaboration, lead

Challenge 2: It is Hard To Control Access

Challenge 2: It is Hard To Control Access

Challenge 3

Polit Behav (2010) 32:369–386
DOI 10.1007/s11109-010-9114-0

ORIGINAL PAPER

Affect, Social Pressure and Prosocial Motivation: Field Experimental Evidence of the Mobilizing Effects of Pride, Shame and Publicizing Voting Behavior

Costas Panagopoulos

Challenge 3

WHO VOTES IS PUBLIC INFORMATION!

Dear registered voter:

On November 6, 2007, an election to select local leaders will be held in Ely, IA.

As a registered voter, you are eligible to vote in this election. We urge you to exercise your civic duty and vote on November 6th.

We also remind you that who votes is a matter of public record.

To promote participation in the election, we will obtain a complete list of registered voters who cast ballots on Election Day from local election officials. Shortly after the November 2007 election, we will publish in the local newspaper a complete list of all Ely registered voters who did <u>not</u> vote.

The names of those who took the time to vote will not appear on this list.

DO YOUR CIVIC DUTY! VOTE ON ELECTION DAY!

Challenge 3: Public/Private Distinction

Challenge 3: Public/Private Distinction

- Conventional public/private dichotomy is complicated in the data age.

Challenge 3: Public/Private Distinction

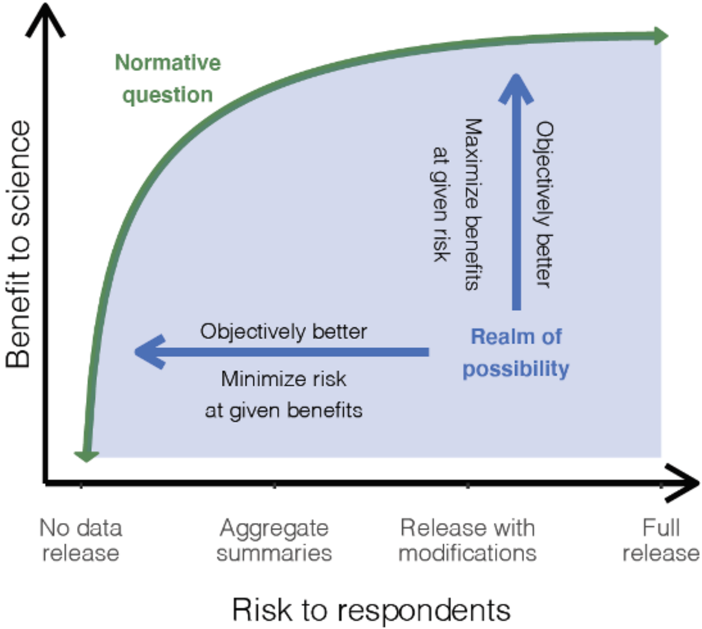
- Conventional public/private dichotomy is complicated in the data age.
- Better to think Hellen Nissenbaum's contextual integrity $\rightsquigarrow$ flows of information.

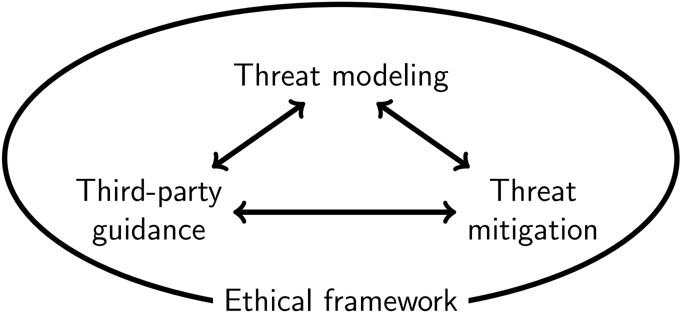
Challenge 3: Public/Private Distinction

- Conventional public/private dichotomy is complicated in the data age.
- Better to think Hellen Nissenbaum's contextual integrity $\rightsquigarrow$ flows of information.
- Key idea is "context-relative informational norms"

Challenge 3: Public/Private Distinction

- Conventional public/private dichotomy is complicated in the data age.
- Better to think Hellen Nissenbaum's contextual integrity $\rightsquigarrow$ flows of information.
- Key idea is "context-relative informational norms"
  - actors (subject, sender, recipient)
  - attributes (types of information)
  - transmission principles (constraints under which information flows)


Normative question
Objectively better
Maximize benefits at given risk
Objectively better
Minimize risk at given benefits
Realm of possibility
Benefit to science
No data release
Aggregate summaries
Release with modifications
Full release
Risk to respondents


Threat modeling
Third-party guidance
Threat mitigation
Ethical framework

Fragile Families
& Child Wellbeing Study
PRINCETON | COLUMBIA

Fragile Families
& Child Wellbeing Study
PRINCETON | COLUMBIA

Features relevant to privacy:

1. Informed consent

Fragile Families
& Child Wellbeing Study
PRINCETON | COLUMBIA

Features relevant to privacy:

1. Informed consent
2. Already available to researchers

Fragile Families
& Child Wellbeing Study
PRINCETON | COLUMBIA

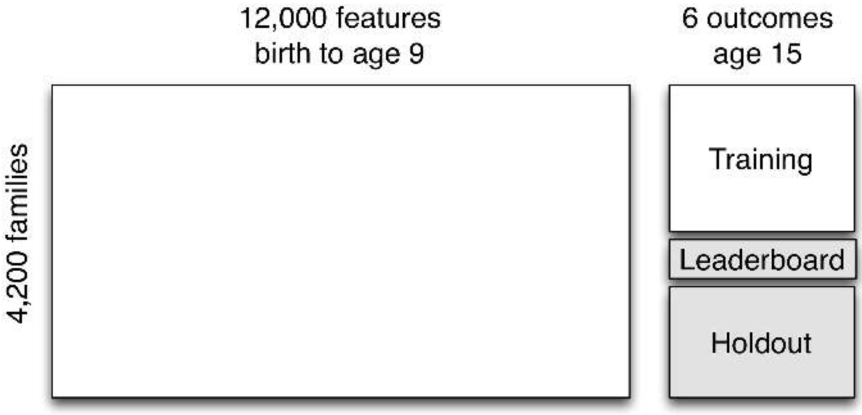
Features relevant to privacy:

1. Informed consent
2. Already available to researchers
3. Already used in scientific and policy debates

Fragile Families
& Child Wellbeing Study
PRINCETON | COLUMBIA

Features relevant to privacy:

1. Informed consent
2. Already available to researchers
3. Already used in scientific and policy debates
4. Contain information from many respondents


12,000 features
birth to age 9
6 outcomes
age 15
4,200 families
Training
Leaderboard
Holdout

The Princeton IRB approved the project.

The Princeton IRB approved the project.

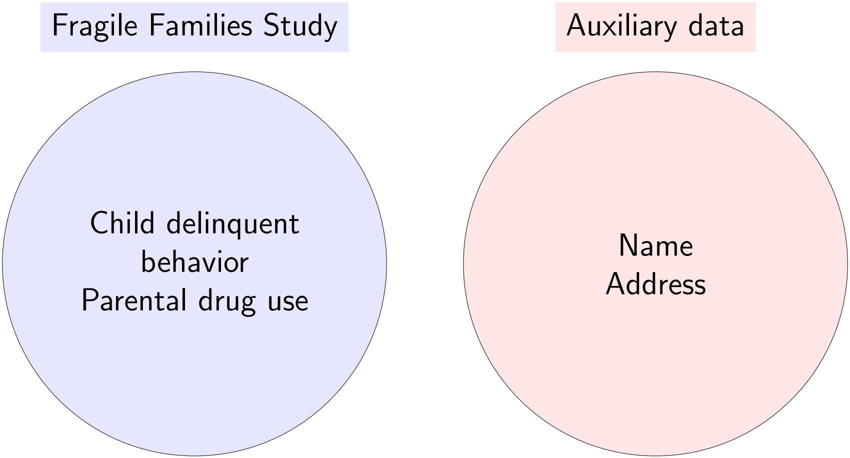
The data met standards for de-identification.

The Princeton IRB approved the project.

The data met standards for de-identification.

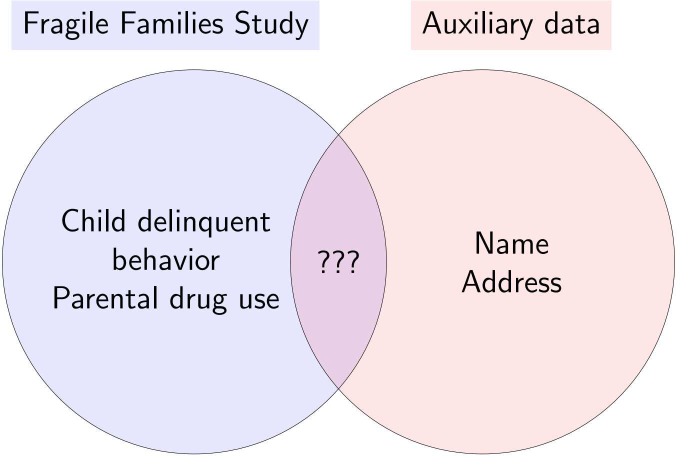
Why worry?

Unknown auxiliary data that may exist.



Slide Credit: Ian Lundberg and Matt Salganik

Unknown auxiliary data that may exist.



Slide Credit: Ian Lundberg and Matt Salganik

Threat modeling

Criteria that represent a threat of a re-identification attack:

- skills
- auxiliary data
- incentives

Threat modeling

Criteria that represent a threat of a re-identification attack:

- skills
- auxiliary data
- incentives

Concrete experience: Attacked their own data for 1.5 months.

Threat modeling

Criteria that represent a threat of a re-identification attack:

- skills
- auxiliary data
- incentives

Concrete experience: Attacked their own data for 1.5 months.

Main threats:

1. Privacy researcher

Threat modeling

Criteria that represent a threat of a re-identification attack:

- skills
- auxiliary data
- incentives

Concrete experience: Attacked their own data for 1.5 months.

Main threats:

1. Privacy researcher
2. Nosy neighbor

Threat modeling

Criteria that represent a threat of a re-identification attack:

- skills
- auxiliary data
- incentives

Concrete experience: Attacked their own data for 1.5 months.

Main threats:

1. Privacy researcher
2. Nosy neighbor
3. Troll

Threat modeling

Criteria that represent a threat of a re-identification attack:

- skills
- auxiliary data
- incentives

Concrete experience: Attacked their own data for 1.5 months.

Main threats:

1. Privacy researcher
2. Nosy neighbor
3. Troll
4. Journalist

Threat modeling

Criteria that represent a threat of a re-identification attack:

- skills
- auxiliary data
- incentives

Concrete experience: Attacked their own data for 1.5 months.

Main threats:

1. Privacy researcher
2. Nosy neighbor
3. Troll
4. Journalist
5. Cheater

← Mitigations →

↑ Threats ↓

	Low profile	Careful language	Challenge structure	Application process	Ethical appeal	Modifications to data
Privacy researcher	✓	✓	✓	✓	✓	✓
Nosy neighbor	✓			✓		
Troll	✓		✓	✓		✓
Journalist	✓	✓	✓	✓	✓	✓
Cheater		✓	✓		✓	✓

Slide Credit: Ian Lundberg and Matt Salganik

Response plan

Mitigated but did not eliminate risks.

Thus needed a team ready to respond in a crisis

Response plan

Mitigated but did not eliminate risks.

Thus needed a team ready to respond in a crisis

- Computer scientist who had re-identified datasets previously
- Lawyer and sociologist who studies privacy and inequality
- Respected journalist

Response plan

Mitigated but did not eliminate risks.

Thus needed a team ready to respond in a crisis

- Computer scientist who had re-identified datasets previously
- Lawyer and sociologist who studies privacy and inequality
- Respected journalist

Prepared to respond **quickly**.

Generalizable Principles

Generalizable Principles

- Do not overstate claims to protect privacy.

Generalizable Principles

- Do not overstate claims to protect privacy.
- Invite adversaries to become collaborators.

Generalizable Principles

- Do not overstate claims to protect privacy.
- Invite adversaries to become collaborators.
- Share data only with those likely to contribute.

Generalizable Principles

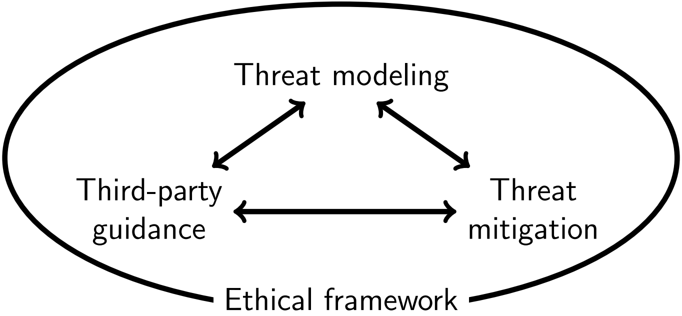
- Do not overstate claims to protect privacy.
- Invite adversaries to become collaborators.
- Share data only with those likely to contribute.
- Embed data sharing within norms of appropriate behavior.

Generalizable Principles

- Do not overstate claims to protect privacy.
- Invite adversaries to become collaborators.
- Share data only with those likely to contribute.
- Embed data sharing within norms of appropriate behavior.
- Only share the data needed for the research project.

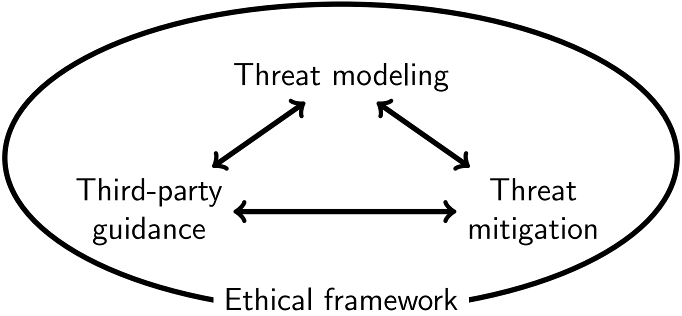
Ethics grounded in the Belmont Report

- Respect for persons
- Beneficence
- Justice



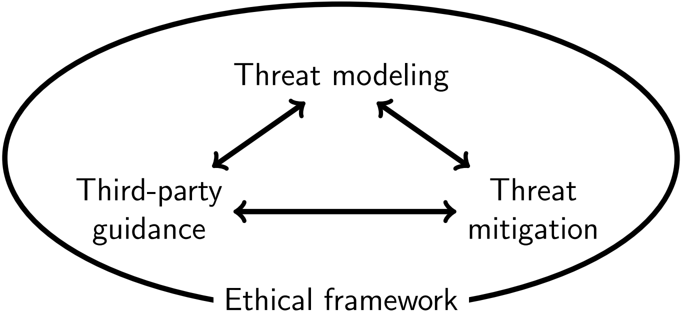
Ethics grounded in the Belmont Report

- Respect for persons → honor participants' agreement
- Beneficence
- Justice



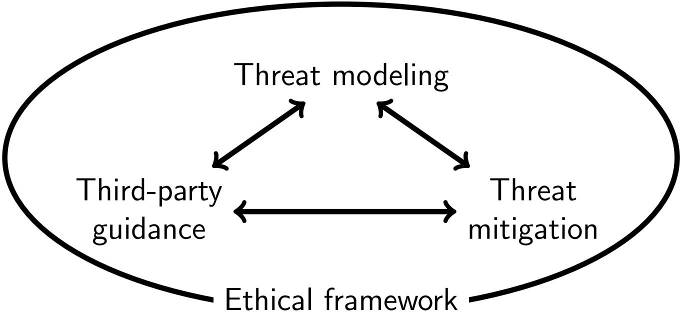
Ethics grounded in the Belmont Report

- Respect for persons → honor participants' agreement
- Beneficence → maximize benefits and minimize harms
- Justice



Ethics grounded in the Belmont Report

- Respect for persons → honor participants' agreement
- Beneficence → maximize benefits and minimize harms
- Justice → population to benefit is similar to study population



This Week in Review

This Week in Review

- Values introduction
- Privacy
- Fragile Families Case Study

This Week in Review

- Values introduction
- Privacy
- Fragile Families Case Study

Next Week: Fairness and Racial Justice

- "Are We Automating Racism?" video
- Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. United Kingdom: Wiley. Introduction–Chapter 2, 5
- Bender, Emily M. and Gebru, Timnit and McMillan-Major, Angelina and Shmitchell, Shmargaret (2021) "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?"