

data-ppf.github.io 2021-04-06

Lec 12: the battle for data “ethics”

chris wiggins + matt jones, Columbia

## student observations: data

223 ethics/ ethical/ ethicality

57 salganik

27 metcalf

21 sweeney

14 principles

3 skeptical

1 pessimistic

## student observations: pessimism

- ▶ As others said before, this week's readings weren't especially hopeful.

## student observations: pessimism

- ▶ As others said before, this week's readings weren't especially hopeful.
- ▶ The Metcalf et.al paper heightened my already-pronounced suspicion of tech companies' internally-directed ethics projects. Of course, there are the typical factors that should make one skeptical of companies professing to have ethical values at heart.

## student observations: pessimism

- ▶ As others said before, this week's readings weren't especially hopeful.
- ▶ The Metcalf et.al paper heightened my already-pronounced suspicion of tech companies' internally-directed ethics projects. Of course, there are the typical factors that should make one skeptical of companies professing to have ethical values at heart.
- ▶ The readings this week made me feel pessimistic about the state of data privacy and ethics.

## student observations: pessimism

- ▶ As others said before, this week's readings weren't especially hopeful.
- ▶ The Metcalf et.al paper heightened my already-pronounced suspicion of tech companies' internally-directed ethics projects. Of course, there are the typical factors that should make one skeptical of companies professing to have ethical values at heart.
- ▶ The readings this week made me feel pessimistic about the state of data privacy and ethics.
- ▶ We could respond with the hopeless feeling that best-intentions are inadequate to preventing unethical behavior, which they are. But I also believe they remain worthwhile

outline for today

# outline for today

1. data and ethics



# outline for today

1. data and ethics
  - 1.1 recall from Neff

# outline for today

1. data and ethics
  - 1.1 recall from Neff
  - 1.2 what went wrong? ethics theater/ethics washing

## outline for today

1. data and ethics
  - 1.1 recall from Neff
  - 1.2 what went wrong? ethics theater/ethics washing
2. “applied ethics”: Belmont in Salganik

# outline for today

1. data and ethics
  - 1.1 recall from Neff
  - 1.2 what went wrong? ethics theater/ethics washing
2. “applied ethics”: Belmont in Salganik
  - 2.1 Belmont

# outline for today

1. data and ethics
  - 1.1 recall from Neff
  - 1.2 what went wrong? ethics theater/ethics washing
2. “applied ethics”: Belmont in Salganik
  - 2.1 Belmont
  - 2.2 principles

# outline for today

1. data and ethics
  - 1.1 recall from Neff
  - 1.2 what went wrong? ethics theater/ethics washing
2. “applied ethics”: Belmont in Salganik
  - 2.1 Belmont
  - 2.2 principles
3. define vs design

# outline for today

1. data and ethics
  - 1.1 recall from Neff
  - 1.2 what went wrong? ethics theater/ethics washing
2. “applied ethics”: Belmont in Salganik
  - 2.1 Belmont
  - 2.2 principles
3. define vs design
  - 3.1 IRB as an example of process design

# outline for today

1. data and ethics
  - 1.1 recall from Neff
  - 1.2 what went wrong? ethics theater/ethics washing
2. “applied ethics”: Belmont in Salganik
  - 2.1 Belmont
  - 2.2 principles
3. define vs design
  - 3.1 IRB as an example of process design
  - 3.2 IRB@FB: from Kramer 2014 to Kanerva 2016



# outline for today

1. data and ethics
  - 1.1 recall from Neff
  - 1.2 what went wrong? ethics theater/ethics washing
2. “applied ethics”: Belmont in Salganik
  - 2.1 Belmont
  - 2.2 principles
3. define vs design
  - 3.1 IRB as an example of process design
  - 3.2 IRB@FB: from Kramer 2014 to Kanerva 2016
  - 3.3 Ethics and people: “Owners” (2019)

# outline for today

1. data and ethics
  - 1.1 recall from Neff
  - 1.2 what went wrong? ethics theater/ethics washing
2. “applied ethics”: Belmont in Salganik
  - 2.1 Belmont
  - 2.2 principles
3. define vs design
  - 3.1 IRB as an example of process design
  - 3.2 IRB@FB: from Kramer 2014 to Kanerva 2016
  - 3.3 Ethics and people: “Owners” (2019)
  - 3.4 Ethics and process: checkpoints, power, theater

# outline for today

1. data and ethics
  - 1.1 recall from Neff
  - 1.2 what went wrong? ethics theater/ethics washing
2. “applied ethics”: Belmont in Salganik
  - 2.1 Belmont
  - 2.2 principles
3. define vs design
  - 3.1 IRB as an example of process design
  - 3.2 IRB@FB: from Kramer 2014 to Kanerva 2016
  - 3.3 Ethics and people: “Owners” (2019)
  - 3.4 Ethics and process: checkpoints, power, theater
4. Ethics and power (more in Lec 14)

# outline for today

1. data and ethics
  - 1.1 recall from Neff
  - 1.2 what went wrong? ethics theater/ethics washing
2. “applied ethics”: Belmont in Salganik
  - 2.1 Belmont
  - 2.2 principles
3. define vs design
  - 3.1 IRB as an example of process design
  - 3.2 IRB@FB: from Kramer 2014 to Kanerva 2016
  - 3.3 Ethics and people: “Owners” (2019)
  - 3.4 Ethics and process: checkpoints, power, theater
4. Ethics and power (more in Lec 14)
  - 4.1 external:

# outline for today

1. data and ethics
  - 1.1 recall from Neff
  - 1.2 what went wrong? ethics theater/ethics washing
2. “applied ethics”: Belmont in Salganik
  - 2.1 Belmont
  - 2.2 principles
3. define vs design
  - 3.1 IRB as an example of process design
  - 3.2 IRB@FB: from Kramer 2014 to Kanerva 2016
  - 3.3 Ethics and people: “Owners” (2019)
  - 3.4 Ethics and process: checkpoints, power, theater
4. Ethics and power (more in Lec 14)
  - 4.1 external:
  - 4.2 internal: people

# outline for today

1. data and ethics
  - 1.1 recall from Neff
  - 1.2 what went wrong? ethics theater/ethics washing
2. “applied ethics”: Belmont in Salganik
  - 2.1 Belmont
  - 2.2 principles
3. define vs design
  - 3.1 IRB as an example of process design
  - 3.2 IRB@FB: from Kramer 2014 to Kanerva 2016
  - 3.3 Ethics and people: “Owners” (2019)
  - 3.4 Ethics and process: checkpoints, power, theater
4. Ethics and power (more in Lec 14)
  - 4.1 external:
  - 4.2 internal: people
  - 4.3 internal: profit (more in Lec 13)

# 1. data and ethics

## 1.1 recall from Neff

- ▶ “How to build ethical data science is far from a solved problem, and data science teams continue to grapple with it”



## 1.1 recall from Neff

- ▶ “How to build ethical data science is far from a solved problem, and data science teams continue to grapple with it”
- ▶ “The data scientists we observed frequently understood and advocated for considering competing values and ethical choices in their work, and were more reflexive about their practices than critical data studies often present them as being.”

## 1.1 recall from Neff

- ▶ “How to build ethical data science is far from a solved problem, and data science teams continue to grapple with it”
- ▶ “The data scientists we observed frequently understood and advocated for considering competing values and ethical choices in their work, and were more reflexive about their practices than critical data studies often present them as being.”
- ▶ “[boyd 2016]: How do we enable ethics in the complex big data systems that are situated within organizations, influenced by diverse intentions and motivations, shaped by politics and organizational logics, complicated by issues of power and control?”

## 1.2 google's answer: AI ethics team



Wired

### [Google Offers to Help Others With the Tricky Ethics of AI](#)

After learning its own ethics lessons the hard way, the tech giant will offer services like spotting racial bias or developing guidelines around AI projects. (170 kB) ▾



Figure 1: Google Aug 2020

## 1.2 google's answer: AI ethics team



Figure 2: Google Dec 2020

## 1.2 google's answer: AI ethics team



U.S. LEGAL NEWS   FEBRUARY 19, 2021 / 5:24 PM / UPDATED A MONTH AGO

# Google fires second AI ethics leader as dispute over research, diversity grows

Figure 3: Google Feb 2021

## 1.3 what went wrong? ethics theater/ethics washing

*I think, in the wake of these controversies, there has been kind of ethics theater, almost. We actually look at this in our 2018 report, where we looked into these a little bit. All of these questions around, "What do these boards actually do," right? Are product decisions run by them? Can they cancel a product decision? Do they have veto power otherwise? Is there any documentation on whether their advice was taken or whether it was not?"*

- ▶ from Meredith Whittaker, note: define vs design

## 1.3 what went wrong? ethics theater/ethics washing

[dataethics.eu](http://dataethics.eu)

1. External Participation: early and regular engagement with all relevant stakeholders.

*Wagner, Ben. "Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?." Being Profiling. Cogitas Ergo Sum (2018).*

## 1.3 what went wrong? ethics theater/ethics washing

[dataethics.eu](http://dataethics.eu)

1. External Participation: early and regular engagement with all relevant stakeholders.
2. Provide a mechanism for external independent oversight.

*Wagner, Ben. "Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?." Being Profiling. Cogitas Ergo Sum (2018).*



## 1.3 what went wrong? ethics theater/ethics washing

[dataethics.eu](http://dataethics.eu)

1. External Participation: early and regular engagement with all relevant stakeholders.
2. Provide a mechanism for external independent oversight.
3. Ensure transparent decision-making procedures on why decisions were taken.

*Wagner, Ben. "Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?." Being Profiling. Cogitas Ergo Sum (2018).*

## 1.3 what went wrong? ethics theater/ethics washing

[dataethics.eu](http://dataethics.eu)

1. External Participation: early and regular engagement with all relevant stakeholders.
2. Provide a mechanism for external independent oversight.
3. Ensure transparent decision-making procedures on why decisions were taken.
4. Develop a stable list of non-arbitrary of standards where the selection of certain values, ethics and rights over others can be plausibly justified.

*Wagner, Ben. "Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?." Being Profiling. Cogitas Ergo Sum (2018).*

## 1.3 what went wrong? ethics theater/ethics washing

[dataethics.eu](http://dataethics.eu)

1. External Participation: early and regular engagement with all relevant stakeholders.
2. Provide a mechanism for external independent oversight.
3. Ensure transparent decision-making procedures on why decisions were taken.
4. Develop a stable list of non-arbitrary of standards where the selection of certain values, ethics and rights over others can be plausibly justified.
5. Ensure that ethics do not substitute fundamental rights or human rights.

*Wagner, Ben. "Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?." Being Profiling. Cogitas Ergo Sum (2018).*

## 1.3 what went wrong? ethics theater/ethics washing

[dataethics.eu](http://dataethics.eu)

1. External Participation: early and regular engagement with all relevant stakeholders.
  2. Provide a mechanism for external independent oversight.
  3. Ensure transparent decision-making procedures on why decisions were taken.
  4. Develop a stable list of non-arbitrary of standards where the selection of certain values, ethics and rights over others can be plausibly justified.
  5. Ensure that ethics do not substitute fundamental rights or human rights.
  6. Provide a clear statement on the relationship between the commitments made and existing legal or regulatory frameworks, in particular on what happens when the two are in conflict.
- Wagner, Ben. "Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?." Being Profiling. Cogitas Ergo Sum (2018).*

what else is there? how can we understand,  
define, design for ethics?

what else is there? how can we understand, define, design for ethics?

1. .com: “What should the doing of ethics look like? An obvious place to turn would be applied business ethics research literatures.” – Metcalf et al

what else is there? how can we understand, define, design for ethics?

1. .com: “What should the doing of ethics look like? An obvious place to turn would be applied business ethics research literatures.” – Metcalf et al
2. .edu/.gov: human subjects research: “rather than treating [critical data studies] as an entirely new field with unique problems”, Metcalf + Crawford 2016 “Where are human subjects in Big Data research?”

## 2. “applied ethics”: Belmont in Salganik



## 2. “applied ethics”: Belmont in Salganik

- ▶ claim: “the principles-based approach is sufficiently general that it will be helpful no matter where you work (e.g., university, government, NGO, or company).”

## 2. “applied ethics”: Belmont in Salganik

- ▶ claim: “the principles-based approach is sufficiently general that it will be helpful no matter where you work (e.g., university, government, NGO, or company).”
- ▶ goal: “to move beyond focusing on what is permitted by existing regulations and increase your ability to communicate your reasoning with other researchers and the public.”

## 2.1 Belmont: context (cw)

## 2.1 Belmont: context

The screenshot shows the front page of The New York Times website from Wednesday, July 26, 1972. The navigation bar includes 'HOME', 'SEARCH ARCHIVE', and 'INDEX'. A 'SUBJECTS' dropdown menu is open, showing a search for 'Syphilis Victims in U.S. Study Went Untreated for 40 Years'. The main article headline is 'Syphilis Victims in U.S. Study Went Untreated for 40 Years'. Below the headline is a sub-headline 'SYPHILIS VICTIMS GOT NO THERAPY'. The article text begins with 'WASHINGTON, July 25 — For 40 years the United States Public Health Service has conducted a study in which human beings with syphilis, who were induced to serve as guinea pigs, have gone without medical treatment for the disease.' A smaller version of the article is visible in the background on the right side of the page.

THE NEW YORK TIMES  
WEDNESDAY, JULY 26, 1972

HOME SEARCH ARCHIVE INDEX

SUBJECTS

Syphilis Victims in U.S. Study Went Untreated for 40 Years

SYPHILIS VICTIMS GOT NO THERAPY

WASHINGTON, July 25 — For 40 years the United States Public Health Service has conducted a study in which human beings with syphilis, who were induced to serve as guinea pigs, have gone without medical treatment for the disease.

Figure 4: your local paper

## 2.1 Belmont: context

**ARCHIVES** | Syphilis Victims in U.S. Study Went Untreated for 40 Years

---

The experiment, called the Tuskegee Study, began in 1932 with about 600 black men mostly poor and uneducated, from Tuskegee, Ala., an area that had the highest syphilis rate in the nation at the time.

Four hundred of the group had syphilis and never received deliberate treatment for the Venereal Infection. A control group of 200 had no syphilis and did not receive any specific therapy.

Some subjects were added to the study In its early years to replace men who had dropped out of the program, but the number added is not known. At the beginning of this year, 74 of those who received no treatment were still alive.

As Incentives to enter the Program, the men were promised free transportation to and from hospitals, free hot lunches, free medicine for any disease other than syphilis and free burial after autopsies were performed.

### **Could Have Been Helped**

The Tuskegee Study began 10 years before penicillin was found to be a cure for syphilis and 15 years before the drug became widely available.

## 2.1 Belmont: context

---

THE EXPERIMENT AND HEW'S ETHICAL REVIEW

---

### Racism and Research: The Case of the Tuskegee Syphilis Study

*by ALLAN M. BRANDT*

---

Figure 6: power, examined

## 2.1 Belmont: context

*ALLAN M. BRANDT is a doctoral candidate in the Department of History, Columbia University. He is presently writing a social history of venereal disease in the United States. Mr. Brandt was a student intern at The Hastings Center in 1977.*

Figure 7: power, examined – grad student

*Amalie Kass Professor of History of Medicine and Professor of ... The Cigarette Century: The Rise, Fall, and Deadly Persistence of the Product that Defined America*

## 2.1 Belmont: context

**ARCHIVES** | Syphilis Victims in U.S. Study Went Untreated for 40 Years

---

### **‘Never Clandestine’**

The syphilis study “was never clandestine” and 15 scientific reports were published in the medical literature, Dr. Millar said in a telephone interview yesterday from Atlanta.

Officials who initiated the study in 1932 had informed the syphilis victims that they could get treatment for the infection at any time, Dr. Millar said.

“Patients were not denied drugs,” Dr. Millar stressed. Rather, they were not offered drugs.

When the study began, doctors could offer only what is now regarded as poor therapy —injections of metals like bismuth, arsenic and mercury. Such treatments were known to be toxic.

Many doctors, Dr. Miller said, then thought “it better not to treat syphilis cases because of the mortality from” the metal therapies.

The critical period in ethics was in the late nineteen lforties and early nineteen–fifties when antibiotics could have been but were not prescribed for the syphilis patients



## 2.1 Belmont: context

- ▶ R. A. Vonderlehr et al., “Untreated Syphilis in the Male Negro: A Comparative Study of Treated and Untreated Cases,” *Journal of the American Medical Association* 107, no. 11 (September 12, 1936): 856, <https://doi.org/10.1001/jama.1936.02770370020006>.

(our italics)

## 2.1 Belmont: context

- ▶ R. A. Vonderlehr et al., “Untreated Syphilis in the Male Negro: A Comparative Study of Treated and Untreated Cases,” Journal of the American Medical Association 107, no. 11 (September 12, 1936): 856, <https://doi.org/10.1001/jama.1936.02770370020006>.
- ▶ A population of untreated individuals from the Tuskegee Alabama area “seemed to offer an *unusual opportunity* to study the untreated syphilitic patient from the beginning of the disease to the death of the infected person.”

(our italics)

## 2.1 Belmont: context

- ▶ R. A. Vonderlehr et al., “Untreated Syphilis in the Male Negro: A Comparative Study of Treated and Untreated Cases,” Journal of the American Medical Association 107, no. 11 (September 12, 1936): 856, <https://doi.org/10.1001/jama.1936.02770370020006>.
- ▶ A population of untreated individuals from the Tuskegee Alabama area “seemed to offer an *unusual opportunity* to study the untreated syphilitic patient from the beginning of the disease to the death of the infected person.”

(our italics)

- ▶ “An opportunity was also offered to compare this process, uninfluenced by modern treatment, with the results” obtained from people who had been treated.

## 2.1 Belmont: brief timeline

- ▶ A1 of NYT 1973-07-26

## 2.1 Belmont: brief timeline

- ▶ A1 of NYT 1973-07-26
- ▶ 1974-05-30: regulations published

## 2.1 Belmont: brief timeline

- ▶ A1 of NYT 1973-07-26
- ▶ 1974-05-30: regulations published
- ▶ 1974-12-03/04: 1st meeting of commission

## 2.1 Belmont: brief timeline

- ▶ A1 of NYT 1973-07-26
- ▶ 1974-05-30: regulations published
- ▶ 1974-12-03/04: 1st meeting of commission
- ▶ 1978-09-08: 43rd (final) meeting of Commission

## 2.1 Belmont: brief timeline

- ▶ A1 of NYT 1973-07-26
- ▶ 1974-05-30: regulations published
- ▶ 1974-12-03/04: 1st meeting of commission
- ▶ 1978-09-08: 43rd (final) meeting of Commission
- ▶ 1979-04-18: federal register, i.e., law



## 2.1 Belmont: goal

- ▶ Beauchamp: “consultant philosopher”  
*“the project of creating a framework of basic principles for all federally funded research had been mandated by a public law enacted by the US congress”*  
*“the national commission viewed these principles as embedded in preexisting public morality”*

## 2.1 Belmont: outputs

- ▶ report, 1979-04-10, 10pp

## 2.1 Belmont: outputs

- ▶ report, 1979-04-10, 10pp
- ▶ recommends IRB, 1978-09-01, 132pp

## 2.1 Belmont: outputs

- ▶ report, 1979-04-10, 10pp
- ▶ recommends IRB, 1978-09-01, 132pp
- ▶ appendix v1 “78-0013”: 611pp

## 2.1 Belmont: outputs

- ▶ report, 1979-04-10, 10pp
- ▶ recommends IRB, 1978-09-01, 132pp
- ▶ appendix v1 “78-0013”: 611pp
- ▶ appendix v2 “78-0014”: 705pp

## 2.2 Belmont: principlism

- ▶ “principles->standards->rules” (con law)

## 2.2 Belmont: principlism

- ▶ “principles->standards->rules” (con law)
- ▶ “comprehensive” – Belmont

## 2.2 Belmont: principlism

- ▶ “principles->standards->rules” (con law)
- ▶ “comprehensive” – Belmont
- ▶ no expectation that principles will not conflict



## 2.2 Belmont: principlism

- ▶ “principles->standards->rules” (con law)
- ▶ “comprehensive” – Belmont
- ▶ no expectation that principles will not conflict
- ▶ tensions will be negotiated and interpreted into rules and process

## 2.2 Belmont: principles

---

Principle

---

Respect for persons

Beneficence

Justice

---

## 2.2 principles and norms

Principle	short
Respect for persons	means
Beneficence	ends
Justice	norms

## 2.2 principles and philosophy

Principle	short	origin
Respect for persons	means	Deontology (Kant)
Beneficence	ends	Consequentialism (Bentham/Mill)
Justice	norms	millennia of law

## 2.2 principles, cartooned

Principle	short	origin	cartoon
Respect for persons	means	Deontology	"informed consent"
Beneficence	ends	Consequentialism	"do no harm"
Justice	norms	millennia of law	"fairness"

## 2.2 principles: what's missed?

*(post in Zoom)*

## 2.2 principles: where's transparency?

- ▶ transparency a “means”

## 2.2 principles: where's transparency?

- ▶ transparency a “means”
  - ▶ respect: informed consent



## 2.2 principles: where's transparency?

- ▶ transparency a “means”
  - ▶ respect: informed consent
  - ▶ beneficence: reveal harms

## 2.2 principles: where's transparency?

- ▶ transparency a “means”
  - ▶ respect: informed consent
  - ▶ beneficence: reveal harms
  - ▶ justice: reveal biases + minoritized risk / “edge cases”

## 2.2 principles: where's privacy?

- ▶ privacy? if defined contextually. . . .

In *Privacy in Context: Technology, Policy, and the Integrity of Social Life*,<sup>7</sup> I give an account of privacy in terms of expected flows of personal information, modeled with the construct of *context-relative informational norms*. The key parameters of informational norms are actors (subject, sender, recipient), attributes (types of information), and transmission principles (constraints under which information flows). Generally, when the flow of information adheres to entrenched norms, all is well; violations of these norms, however, often result in protest and complaint. In a health care context, for example, patients expect their physicians to keep personal medical information confidential, yet they accept that it might be shared with specialists as needed. Patients' expectations would be breached and they would likely be shocked and dismayed if they learned that their physicians had sold the information to a marketing company. In this event, we would say that informational norms for the health care context had been violated.

Figure 8: Helen Nissenbaum

reference: Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (2010).

from "Nissenbaum, Helen." A contextual approach to privacy online." *Daedalus* 140, no. 4 (2011): 32-48."

## 2.2 principles: where's privacy?

- ▶ privacy? if defined contextually. . . .

In *Privacy in Context: Technology, Policy, and the Integrity of Social Life*,<sup>7</sup> I give an account of privacy in terms of expected flows of personal information, modeled with the construct of *context-relative informational norms*. The key parameters of informational norms are actors (subject, sender, recipient), attributes (types of information), and transmission principles (constraints under which information flows). Generally, when the flow of information adheres to entrenched norms, all is well; violations of these norms, however, often result in protest and complaint. In a health care context, for example, patients expect their physicians to keep personal medical information confidential, yet they accept that it might be shared with specialists as needed. Patients' expectations would be breached and they would likely be shocked and dismayed if they learned that their physicians had sold the information to a marketing company. In this event, we would say that informational norms for the health care context had been violated.

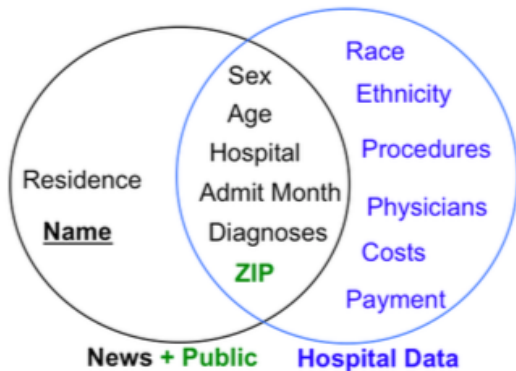
Figure 8: Helen Nissenbaum

reference: Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (2010).

from "Nissenbaum, Helen." A contextual approach to privacy online." *Daedalus* 140, no. 4 (2011): 32-48."

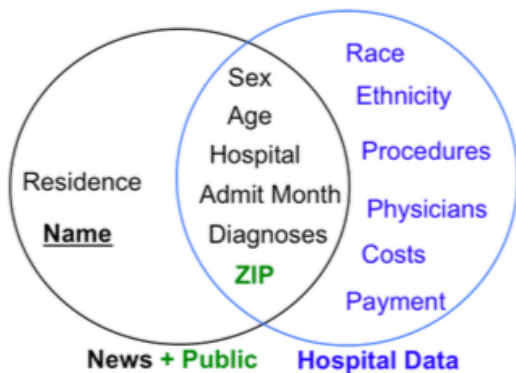
## 2.2 principles, privacy, and “harms”

- ▶ mitigating data harms: not a trolley problem



## 2.2 principles, privacy, and “harms”

- ▶ mitigating data harms: not a trolley problem
- ▶ almost any data could be a database of ruin



## 2.2 principles: where's accountability

(cf. original “FATML” then “FAT-star” conferences, later FAccT)

- ▶ accountability? wait that's about *power*, which leads to

### 3. define vs design



## 3.1 IRB as an example of process design

- ▶ where is power located?

## 3.1 IRB as an example of process design

- ▶ where is power located?
- ▶ when is power exercised?

## 3.2 IRB@FB: from Kramer 2014 to Kanerva 2016

### Washington and Lee Law Review Online

---

Volume 72 | Issue 3

---

6-14-2016

## Evolving the IRB: Building Robust Review for Industry Research

Molly Jackman

*Facebook*

Lauri Kanerva

*Facebook*

## 3.2 Kanerva:

- ▶ context (Kramer et al 2014)

## 3.2 Kanerva:

- ▶ context (Kramer et al 2014)
- ▶ limits

## 3.2 Kanerva:

- ▶ context (Kramer et al 2014)
- ▶ limits
  - ▶ to be fair, this is for “Industry Research”, not for operations+product

### 3.3 Ethics and people: “Owners” (2019)

*ethics is in tension with meritocracy, technological solutionism, and market fundamentalism. . . . ethics can work to uphold and affirm those other logics.*

### 3.3 Ethics and people: “Owners” (2019)

*D'Ignazio and Klein (2019) argue that the very use of the term “ethics” by the tech sector upholds what Ruha Benjamin (2019) describes as “imagined objectivity” because this term “locate[s] the source of the problem in individuals or technical systems.” . . . “This presents a new problem for sociotechnical scholars used to being ignored: What if, instead of being brushed aside, our critiques are being heard but transformed into something we might not recognize?”*



## 3.4 Ethics and process:

## 3.4 Ethics and process: audits

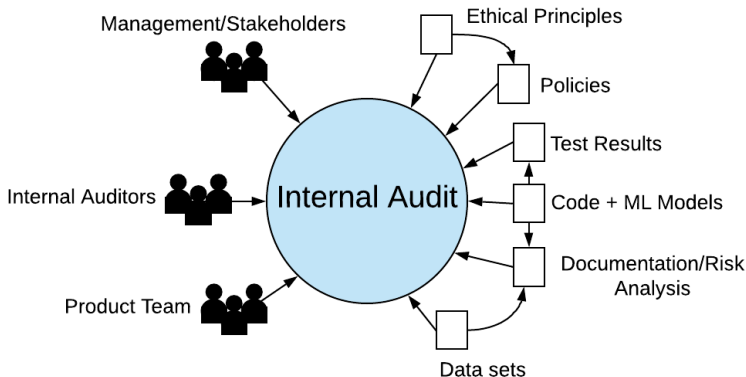
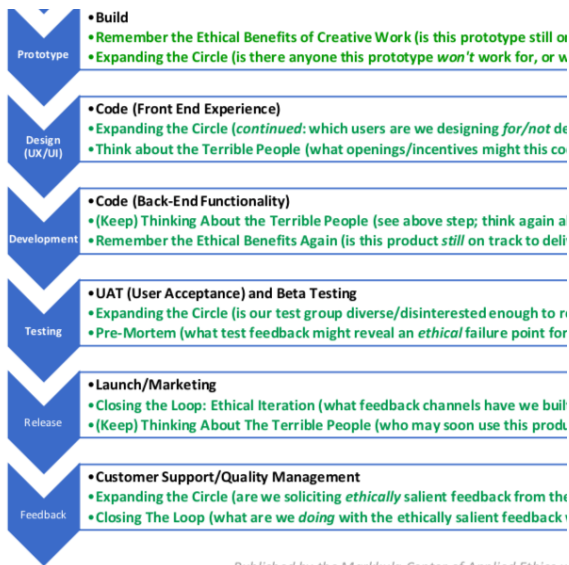


Figure 9: Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, Parker Barnes

## 3.4 Ethics and process: checkpoints



Published by the Markkula Center of Applied Ethics at Santa Clara University

### 3.5 Ethics and process: checkpoints, power, theater (reprise)

*I think, in the wake of these controversies, there has been kind of ethics theater, almost. We actually look at this in our 2018 report, where we looked into these a little bit. All of these questions around, "What do these boards actually do," right? Are product decisions run by them? Can they cancel a product decision? Do they have veto power otherwise? Is there any documentation on whether their advice was taken or whether it was not?"*

- ▶ from Meredith Whittaker, note: define vs design

## 4. Ethics and power

## 4. Ethics and power

- ▶ we can define and design, but how to we give ethics teeth?

## 4.1 external: law, esp. “Human Rights”

*At the 2018 AI Now conference’s public symposium, human rights legal scholar Phillip Alston half-jokingly said from the stage, “I want to strangle ethics,” noting that ethics is “open-ended,” “undefined and unaccountable” in comparison to human and social rights frameworks (Alston 2018).*

- ▶ law is “a floor not a ceiling”

## 4.1 external: law, esp. “Human Rights”

*At the 2018 AI Now conference's public symposium, human rights legal scholar Phillip Alston half-jokingly said from the stage, “I want to strangle ethics,” noting that ethics is “open-ended,” “undefined and unaccountable” in comparison to human and social rights frameworks (Alston 2018).*

- ▶ law is “a floor not a ceiling”
- ▶ unenforced in US, no ICC



## 4.2 internal: people, from Metcalf

- ▶ “To better understand these pitfalls, we approach ethics and morality as social phenomena and not as primarily philosophical abstractions.” – Metcalf et al.

## 4.2 internal: people, from Metcalf

- ▶ “To better understand these pitfalls, we approach ethics and morality as social phenomena and not as primarily philosophical abstractions.” – Metcalf et al.
- ▶ “the system that you create has to be something that people feel adds value and is not a massive roadblock that adds no value, because if it is a roadblock that has no value, people literally won't do it, because they don't have to.” (informant in Metcalf)

## 4.2 internal: people, from Metcalf

- ▶ “To better understand these pitfalls, we approach ethics and morality as social phenomena and not as primarily philosophical abstractions.” – Metcalf et al.
- ▶ “the system that you create has to be something that people feel adds value and is not a massive roadblock that adds no value, because if it is a roadblock that has no value, people literally won’t do it, because they don’t have to.” (informant in Metcalf)
- ▶ Business ethics should include “institutionalizing ethics within organizations as the mutual understanding and alignment of organizational expectations and individual behavior.”

## 4.3 internal: profit

- ▶ ethics must hold power to account, how is this in conflict?

## 4.3 internal: profit

- ▶ ethics must hold power to account, how is this in conflict?
  - ▶ general biz ethics: money vs risk

## 4.3 internal: profit

- ▶ ethics must hold power to account, how is this in conflict?
  - ▶ general biz ethics: money vs risk
  - ▶ b2c ethics: consumer harms vs consumer profits

## 4.3 internal: profit

- ▶ ethics must hold power to account, how is this in conflict?
  - ▶ general biz ethics: money vs risk
  - ▶ b2c ethics: consumer harms vs consumer profits
  - ▶ in ad model, privacy is inconvenient: next week

## 4.3 internal: profit

- ▶ ethics must hold power to account, how is this in conflict?
  - ▶ general biz ethics: money vs risk
  - ▶ b2c ethics: consumer harms vs consumer profits
  - ▶ in ad model, privacy is inconvenient: next week
  - ▶ in ad model, consumer as product: next week



## appendix

## appendix

- ▶ 2021-01-12: intro to course

## appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes

## appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics

## appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism

## appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy

## appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism

## appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation



## appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation
- ▶ 2021-03-09: birth and death of AI

## appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation
- ▶ 2021-03-09: birth and death of AI
- ▶ 2021-03-16: big data, old school (1958-1980)

## appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation
- ▶ 2021-03-09: birth and death of AI
- ▶ 2021-03-16: big data, old school (1958-1980)
- ▶ 2021-03-23: AI2.0

## appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation
- ▶ 2021-03-09: birth and death of AI
- ▶ 2021-03-16: big data, old school (1958-1980)
- ▶ 2021-03-23: AI2.0
- ▶ 2021-03-30: data science, 1962-2017

## appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation
- ▶ 2021-03-09: birth and death of AI
- ▶ 2021-03-16: big data, old school (1958-1980)
- ▶ 2021-03-23: AI2.0
- ▶ 2021-03-30: data science, 1962-2017
- ▶ 2021-04-06: ethics

## appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation
- ▶ 2021-03-09: birth and death of AI
- ▶ 2021-03-16: big data, old school (1958-1980)
- ▶ 2021-03-23: AI2.0
- ▶ 2021-03-30: data science, 1962-2017
- ▶ 2021-04-06: ethics
- ▶ 2021-04-13: present problems: attention economy+VC=dumpsterfire

## appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation
- ▶ 2021-03-09: birth and death of AI
- ▶ 2021-03-16: big data, old school (1958-1980)
- ▶ 2021-03-23: AI2.0
- ▶ 2021-03-30: data science, 1962-2017
- ▶ 2021-04-06: ethics
- ▶ 2021-04-13: present problems: attention economy+VC=dumpsterfire
- ▶ 2021-04-15: future solutions