data-ppf.github.io 2020-04-14

lecture 12 of 14: 'what we talk about when we talk about ethics'

chris wiggins + matt jones, Columbia

student observations: data

```
...
10 facebook
8 hipaa
7 panopticon
5 pandemic
5 organizational
5 $50
4 coronavirus
4 bluetooth
3 virus
...
```

student observations: power

the seemingly only consequence for these serious ethics breaches was public backlash

Salganik's piece and the three examples discussed reveal the power of public backlash in the face of lost privacy,

In this way, relying on researchers to integrate Salganik's "Common Rule" might relate back to Metcalf and Moss's analysis of the factors that oppose more ethical standards in corporations collecting data, such as money/corporate interest, or even the simplicity of not acting ethically when it is unclear if you will be caught

I'm skeptical about the survivability of this ethics trend - whether it can actually achieve the ideal in which ethical violations are a thing of the past

student observations: our valuable data

I started this week's readings with Sweeney's piece and I was honestly a little shocked I knew companies were involved in the selling of user data and such, but I never realized the government encouraged (and sometimes required it)

On the other hand, there are benefits of accessibility to medical history and data, including research in not only the medical field, but also about inequality in the US (who gets what treatment and where, etc)

Both the Salganik and Metcalf et al article brought up ideas I had not considered about ethics in data science, with the first, from the Sagalnik piece, being the idea that privacy concerns could be preempting valuable research from occurring

student observations: principles

The Salganik chapter argues for the use of principles-based approach to ethics that is transferable to all sectors, rather than just one,

Metcalf et al point out that the work of ethics consists of working through and not resolving these tensions of underlying logics and normative goals.

It's a little funny to me that the report includes Respect for Law and Public Interest, when sometimes laws (or lack thereof) are inherently biased themselves and don't protect all people equally

themes for today (1970-present)

themes for today (1970-present)

- reminder: last week: ML ("AI2.0"), and its impact

themes for today (1970-present)

- reminder: last week: ML ("AI2.0"), and its impact
  - the tech

themes for today (1970-present)

- reminder: last week: ML ("AI2.0"), and its impact
  - the tech
  - interpretability/fairness

themes for today (1970-present)

- reminder: last week: ML ("AI2.0"), and its impact
  - the tech
  - interpretability/fairness
  - role of ethics/privacy

themes for today (1970-present)

- reminder: last week: ML ("AI2.0"), and its impact
  - the tech
  - interpretability/fairness
  - role of ethics/privacy
- today: how researchers *define* ethics (1974-1979)

themes for today (1970-present)

- reminder: last week: ML ("AI2.0"), and its impact
  - the tech
  - interpretability/fairness
  - role of ethics/privacy
- today: how researchers *define* ethics (1974-1979)
  - areas missed by the pre-digital, 1970s approach

themes for today (1970-present)

- reminder: last week: ML ("AI2.0"), and its impact
  - the tech
  - interpretability/fairness
  - role of ethics/privacy
- today: how researchers *define* ethics (1974-1979)
  - areas missed by the pre-digital, 1970s approach
- today: how researchers *design for* ethics

themes for today (1970-present)

- reminder: last week: ML ("AI2.0"), and its impact
  - the tech
  - interpretability/fairness
  - role of ethics/privacy
- today: how researchers *define* ethics (1974-1979)
  - areas missed by the pre-digital, 1970s approach
- today: how researchers *design for* ethics
  - remaining 2 weeks: political dynamics of data ethics IRL

readings: Salganik, Sweeney, Metcalf

Salganik, esp. Belmont

- Belmont 1 of many ethics

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

Salganik, esp. Belmont

- Belmont 1 of many ethics
- they worked *so hard*

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

Salganik, esp. Belmont

- Belmont 1 of many ethics
- they worked *so hard*
- influenced *so much* (all researchers, even FB)

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

Salganik, esp. Belmont

- Belmont 1 of many ethics
- they worked *so hard*
- influenced *so much* (all researchers, even FB)
- has been *so stress-tested* (abundant literature)

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

“the Road to Belmont”: Tuskegee as flash point

"the Road to Belmont": Tuskegee as flash point

Figure 1: your local paper

Tuskegee and race

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.

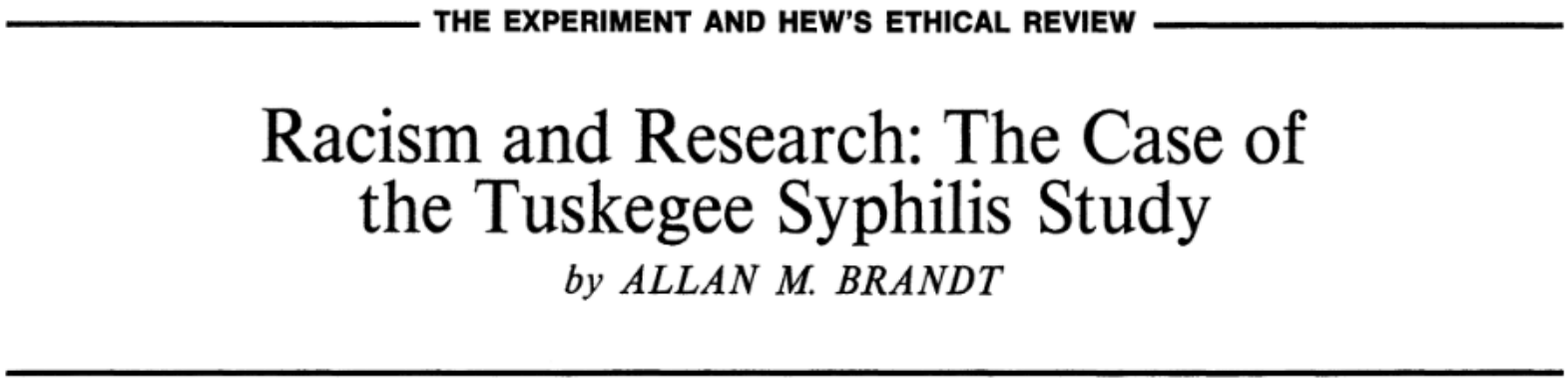
Some subjects were added to the study In its early years to replace men who had dropped out of the program, but the number added is not known. At the beginning of this year, 74 of those who received no treatment were still alive.

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.

Tuskegee and race, 2



THE EXPERIMENT AND HEW'S ETHICAL REVIEW

Racism and Research: The Case of the Tuskegee Syphilis Study

by ALLAN M. BRANDT

Figure 3: power, examined

Tuskegee and race, 3

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 4: 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

Tuskegee: 'never clandestine'

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

Tuskegee: in full view

The American Medical Association meeting May 1936 a lecture on untreated syphilis in black American men.

> *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.*

The results were clear: treatment has dramatically positive effects. AND YET ....

- 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.

“the Road to Belmont”

- Tuskegee as flash point

“the Road to Belmont”

- Tuskegee as flash point
  - fed $

“the Road to Belmont”

- Tuskegee as flash point
  - fed $
  - all the horror and history of American race+power+science in 1 headline

“the Road to Belmont”

- Tuskegee as flash point
  - fed $
  - all the horror and history of American race+power+science in 1 headline
- A1 of NYT 1973-07-26

“the Road to Belmont”

- Tuskegee as flash point
  - fed $
  - all the horror and history of American race+power+science in 1 headline
- A1 of NYT 1973-07-26
- 1974-05-30: regulations published

“the Road to Belmont”

- Tuskegee as flash point
  - fed $
  - all the horror and history of American race+power+science in 1 headline
- A1 of NYT 1973-07-26
- 1974-05-30: regulations published
- 1974-12-03/04: 1st meeting of commission

“the Road to Belmont”

- Tuskegee as flash point
  - fed $
  - all the horror and history of American race+power+science in 1 headline
- 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

“the Road to Belmont”

- Tuskegee as flash point
  - fed $
  - all the horror and history of American race+power+science in 1 headline
- 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

Belmont report

- 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 pubic morality”

Belmont: outputs

- report, 1979-04-10, 10pp

Belmont: outputs

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

Belmont: outputs

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

Belmont: outputs

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

Principlism: granularity

- “frameworks->principles->rules” (Salg)

Principlism: granularity

- “frameworks->principles->rules” (Salg)
- “principles->standards->rules” (con law)

Principlism: granularity

- “frameworks->principles->rules” (Salg)
- “principles->standards->rules” (con law)
- no expectation that principles will not conflict

Principlism: granularity

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

Principlism: tension

- "clarifies trade-offs. . . striking balance"

Principlism: tension

- "clarifies trade-offs. . . striking balance"
- ethical uncertainty prevents research

Principlism: tension

- "clarifies trade-offs. . . striking balance"
- ethical uncertainty prevents research
  - Ebola 2014

Principlism: tension

- "clarifies trade-offs. . . striking balance"
- ethical uncertainty prevents research
  - Ebola 2014
  - Sweeney's work

Principles

- Belmont

Principles

- Belmont
  - person-hood

Principles

- Belmont
  - person-hood
    - informed consent, thus transparency

Principles

- Belmont
  - person-hood
    - informed consent, thus transparency
  - beneficence

Principles

- Belmont
  - person-hood
    - informed consent, thus transparency
  - beneficence
    - do no harm: not just to participants, e.g., society

Principles

- Belmont
  - person-hood
    - informed consent, thus transparency
  - beneficence
    - do no harm: not just to participants, e.g., society
    - "broader impact" (e.g., reproducibility)

Principles

- Belmont
  - person-hood
    - informed consent, thus transparency
  - beneficence
    - do no harm: not just to participants, e.g., society
    - "broader impact" (e.g., reproducibility)
  - justice

Principles

- Belmont
  - person-hood
    - informed consent, thus transparency
  - beneficence
    - do no harm: not just to participants, e.g., society
    - "broader impact" (e.g., reproducibility)
  - justice
    - fairness

Belmont: beneficence vs. personhood

- inadequacy of utilitarian calculations of public good

Certain groups, such as racial minorities, the economically disadvantaged, the very sick, and the institutionalized may continually be sought as research subjects, owing to their ready availability in settings where research is conducted. Given their dependent status and their frequently compromised capacity for free consent, they should be protected against the danger of being involved in research solely for administrative convenience, or because they are easy to manipulate as a result of their illness or socioeconomic condition.

Belmont: beneficence vs. personhood

- inadequacy of utilitarian calculations of public good
- need always to focus on less powerful groups

  Certain groups, such as racial minorities, the economically disadvantaged, the very sick, and the institutionalized may continually be sought as research subjects, owing to their ready availability in settings where research is conducted. Given their dependent status and their frequently compromised capacity for free consent, they should be protected against the danger of being involved in research solely for administrative convenience, or because they are easy to manipulate as a result of their illness or socioeconomic condition.

Belmont: beneficence vs. personhood

- inadequacy of utilitarian calculations of public good
- need always to focus on less powerful groups

  Certain groups, such as racial minorities, the economically disadvantaged, the very sick, and the institutionalized may continually be sought as research subjects, owing to their ready availability in settings where research is conducted. Given their dependent status and their frequently compromised capacity for free consent, they should be protected against the danger of being involved in research solely for administrative convenience, or because they are easy to manipulate as a result of their illness or socioeconomic condition.

- https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html

Has Belmont “kept up with the times”?

- Privacy

Has Belmont “kept up with the times”?

- Privacy
- Informed consent 2.0

Has Belmont “kept up with the times”?

- Privacy
- Informed consent 2.0
  - see your local paper

Principles other than Belmont

- Menlo addendum

  sometimes laws (or lack thereof) are inherently biased themselves and don't protect all people equally.

Principles other than Belmont

- Menlo addendum

  sometimes laws (or lack thereof) are inherently biased themselves and don't protect all people equally.

- Beauchamp

Principles other than Belmont

- Menlo addendum

  sometimes laws (or lack thereof) are inherently biased themselves and don't protect all people equally.

- Beauchamp
- Google AI

Principles other than Belmont

- Menlo addendum

  sometimes laws (or lack thereof) are inherently biased themselves and don't protect all people equally.

- Beauchamp
- Google AI
- FAT/FATE/FATESS

Principles other than Belmont

- Menlo addendum

  sometimes laws (or lack thereof) are inherently biased themselves and don't protect all people equally.

- Beauchamp
- Google AI
- FAT/FATE/FATESS
- ...

Alternative Ethical Systems

- normative ethics

Alternative Ethical Systems

- normative ethics
- virtue ethics

Alternative Ethical Systems

- normative ethics
- virtue ethics
- parables

Alternative Ethical Systems

- normative ethics
- virtue ethics
- parables
- checklists

Alternative Ethical Systems

- normative ethics
- virtue ethics
- parables
- checklists
- utilitarian

Alternative Ethical Systems

- normative ethics
- virtue ethics
- parables
- checklists
- utilitarian
- Kantian

Alternative Ethical Systems

- normative ethics
- virtue ethics
- parables
- checklists
- utilitarian
- Kantian
- casuistry

Alternative Ethical Systems

- normative ethics
- virtue ethics
- parables
- checklists
- utilitarian
- Kantian
- casuistry
- ....

Alternative Ethical Philosophies

- enlightenment: "ideal of progress"

> *accept that we live in a technological culture in which the constant introduction and utilization of new technologies is a normal part of how society works. It then asks how we can deal with such new technologies in a responsible manner,*

Alternative Ethical Philosophies

- enlightenment: "ideal of progress"
- 1970s-1980s: "applied ethics" and, separately "professional ethics"

accept that we live in a technological culture in which the constant introduction and utilization of new technologies is a normal part of how society works. It then asks how we can deal with such new technologies in a responsible manner,

Alternative Ethical Philosophies

- enlightenment: "ideal of progress"
- 1970s-1980s: "applied ethics" and, separately "professional ethics"
  - "largely independently from mainstream philosophy of technology"

accept that we live in a technological culture in which the constant introduction and utilization of new technologies is a normal part of how society works. It then asks how we can deal with such new technologies in a responsible manner,

Alternative Ethical Philosophies

- enlightenment: "ideal of progress"
- 1970s-1980s: "applied ethics" and, separately "professional ethics"
  - "largely independently from mainstream philosophy of technology"
- more recent: "design turn"

  accept that we live in a technological culture in which the constant introduction and utilization of new technologies is a normal part of how society works. It then asks how we can deal with such new technologies in a responsible manner,

Alternative Ethical Philosophies

- enlightenment: "ideal of progress"
- 1970s-1980s: "applied ethics" and, separately "professional ethics"
  - "largely independently from mainstream philosophy of technology"
- more recent: "design turn"

  accept that we live in a technological culture in which the constant introduction and utilization of new technologies is a normal part of how society works. It then asks how we can deal with such new technologies in a responsible manner,

- Brey, "Philosophy of technology after the empirical turn." Techné: Research in Philosophy and Technology (2010)

Alternative Ethical Philosophies

- enlightenment: "ideal of progress"
- 1970s-1980s: "applied ethics" and, separately "professional ethics"
  - "largely independently from mainstream philosophy of technology"
- more recent: "design turn"

  accept that we live in a technological culture in which the constant introduction and utilization of new technologies is a normal part of how society works. It then asks how we can deal with such new technologies in a responsible manner,

- Brey, "Philosophy of technology after the empirical turn." Techné: Research in Philosophy and Technology (2010)
- Van den Hoven, et al. "The design turn in applied ethics." Designing in ethics (2017)

Ethics: define vs design (intra mūrōs)

- ‘accountability’

Ethics: define vs design (intra mūrōs)

- 'accountability'
- IRB "most [scientists] experience ethical debate. . . through the bureaucratic process of IRB"

Ethics: define vs design (intra mūrōs)

- 'accountability'
- IRB "most [scientists] experience ethical debate. . . through the bureaucratic process of IRB"
  - why limit researchers? "IRBs in the United States are required to include at least one non-researcher."

researchers Q:

> *so long as most of these principles rely on the interpretation of researchers, making them prone to biases, we have a long way to go*

–Bennett Brown

Belmont A:

> *The Commission's deliberations on Institutional Review Boards began with the premise that investigators should not have sole responsibility for determining whether research involving human subjects fulfills ethical standards. Others who are independent of the research must share this responsibility, because investigators have a potential conflict by virtue of their concern with the pursuit of knowledge as well as the welfare of the human subjects of their research.*

1978-09-01 IRB recommendation

Ethics: define vs design (extra mūrōs)

- ORB (Jackman + Kanerva 2016)

Ethics: define vs design (extra mūrōs)

- ORB (Jackman + Kanerva 2016)
- what else is there?

Ethics: define vs design (extra mūrōs)

- ORB (Jackman + Kanerva 2016)
- what else is there?
  - design as ethics

Ethics: define vs design (extra mūrōs)

- ORB (Jackman + Kanerva 2016)
- what else is there?
  - design as ethics
  - agile software as ethics

Ethics: define vs design, and principles

INTERVIEWER: You have to feel good about it today.

DR. BEAUCHAMP: Well, yeah, it worked out well, exactly.

INTERVIEWER: Did you expect that the report would be translated as the basis for regulations to the extent that it was?

DR. BEAUCHAMP: No, it is a surprise to me, today, that that happened and that it continues to sustain the kind of interest that it sustained.

Sweeney

Bio on Sweeney

- PhD in CS, MIT w/Hal Abelson

Bio on Sweeney

- PhD in CS, MIT w/Hal Abelson
- prior work:

Bio on Sweeney

- PhD in CS, MIT w/Hal Abelson
- prior work:
  - "arrest records" & name search results 2013

Bio on Sweeney

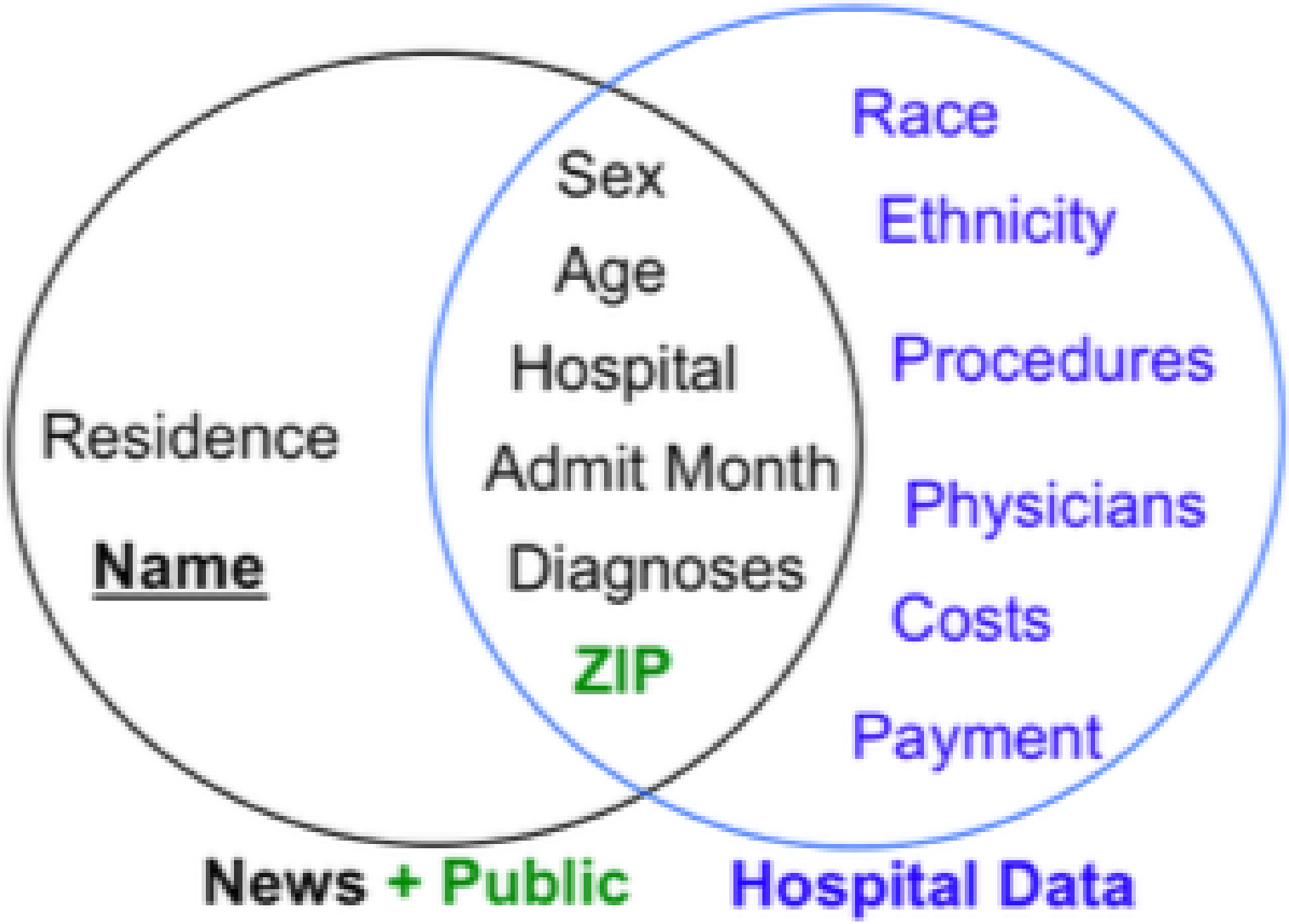
- PhD in CS, MIT w/Hal Abelson
- prior work:
  - "arrest records" & name search results 2013
  - more than I can list. . . .

Sweeney insight on code and law

"The design of the technology, and how it works is really the new policy. And the thing about these designs as policymakers is that we didn't vote for them, we didn't elect them, and we didn't have any say in the things that they believed in. And yet, the decisions that they make turn out to be the rules that we have to live by."

- FAT-star conference 2018, NYU

Sweeney method



Key technical idea:

- public columns as composite key

Key technical idea:

- public columns as composite key
- tech way out: k-anonymity

Key technical idea:

- public columns as composite key
- tech way out: k-anonymity
- mentioned in passing: differential privacy (2006, MSR)

Key technical idea:

- public columns as composite key
- tech way out: k-anonymity
- mentioned in passing: differential privacy (2006, MSR)
  - see also article in your local paper by your local statistician j-school prof/Brown institute director

Key technical idea:

- public columns as composite key
- tech way out: k-anonymity
- mentioned in passing: differential privacy (2006, MSR)
  - see also article in your local paper by your local statistician j-school prof/Brown institute director
- related: what even is privacy?

What even is privacy?

In *Privacy in Context: Technology, Policy, and the Integrity of Social Life,*[7] 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 6: 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."

Beneficence & disclosure, 1/3

While the court praised my skill and advocates dubbed me "the goddess of re-identification," the court ordered knowledge of my method sealed, barring me from publication. Similar fates awaited my other early re-identifications of survey and pharmaceutical data.

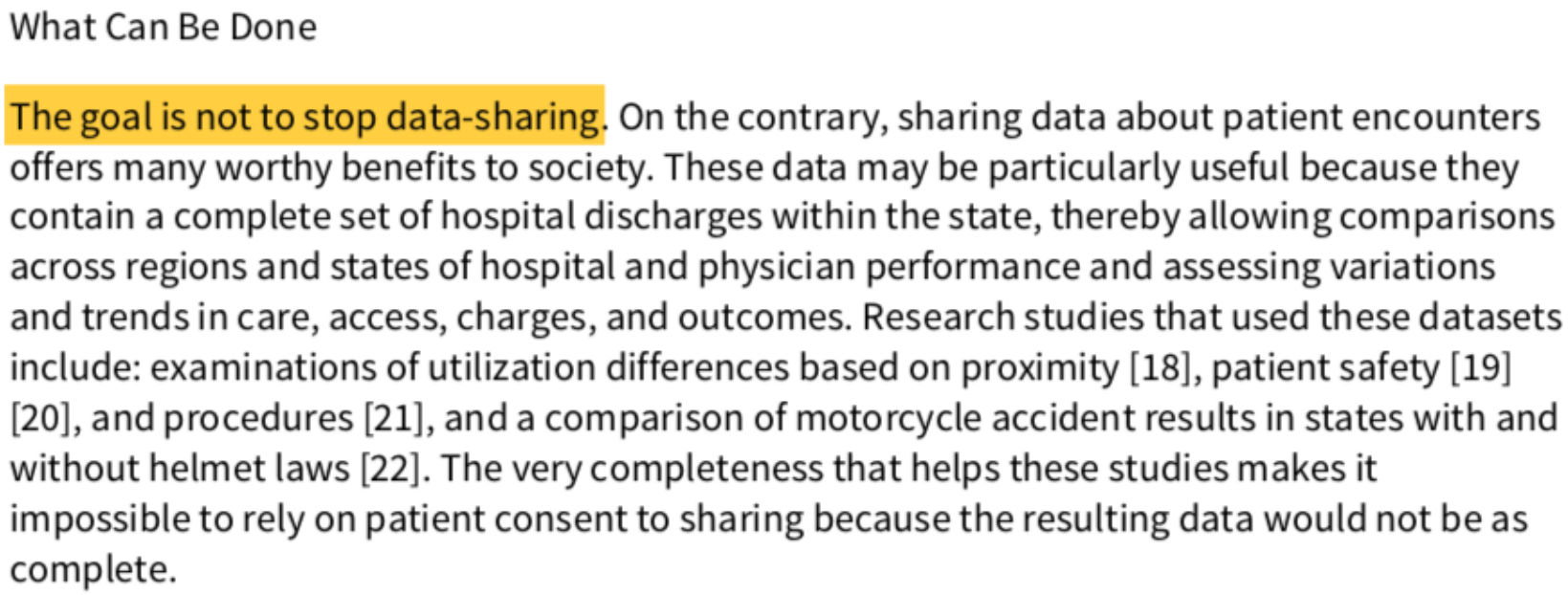
Beneficence & disclosure, 2/3

Computer-science publications refused to publish re-identification experiments unless the paper also included a technological solution, notwithstanding assertions that publishing these experiments would inspire technological innovation to address the real-world problem.

Beneficence & disclosure, 3/3

funding sources refused to fund re-identification experiments unless there was a promise that results would likely show that no risk existed or that all problems could be solved by some promising new theoretical technology under development. Financial resources were unavailable to support rigorous scientific studies otherwise.

Sweeney: response 1/2



What Can Be Done

The goal is not to stop data-sharing. On the contrary, sharing data about patient encounters offers many worthy benefits to society. These data may be particularly useful because they contain a complete set of hospital discharges within the state, thereby allowing comparisons across regions and states of hospital and physician performance and assessing variations and trends in care, access, charges, and outcomes. Research studies that used these datasets include: examinations of utilization differences based on proximity [18], patient safety [19] [20], and procedures [21], and a comparison of motorcycle accident results in states with and without helmet laws [22]. The very completeness that helps these studies makes it impossible to rely on patient consent to sharing because the resulting data would not be as complete.

Figure 7: reacts, 1/2

Sweeney: response 2/2

What Can Be Done

The goal is not to stop data-sharing. On the contrary, sharing data about patient encounters offers many worthy benefits to society. These data may be particularly useful because they contain a complete set of hospital discharges within the state, thereby allowing comparisons across regions and states of hospital and physician performance and assessing variations and trends in care, access, charges, and outcomes. Research studies that used these datasets include: examinations of utilization differences based on proximity [18], patient safety [19] [20], and procedures [21], and a comparison of motorcycle accident results in states with and without helmet laws [22]. The very completeness that helps these studies makes it impossible to rely on patient consent to sharing because the resulting data would not be as complete.

Another goal is to be smarter about how we perform data sharing. This is particularly important as the top buyers of statewide databases are not researchers but private companies, especially those constructing data profiles on individuals [23].

Figure 8: reacts, 2/2

Owning ethics: Logics

> *Ethics is arguably the hottest product in Silicon Valley's hype cycle today"*

1. Meritocracy: "post hoc justification for inequality in society"

Owning ethics: Logics

> *Ethics is arguably the hottest product in Silicon Valley's hype cycle today"*

1. Meritocracy: "post hoc justification for inequality in society"
2. Technological Solutionism

Owning ethics: Logics

> *Ethics is arguably the hottest product in Silicon Valley's hype cycle today"*

1. Meritocracy: "post hoc justification for inequality in society"
2. Technological Solutionism
3. Market Fundamentalism

Owning ethics: Pitfalls

1. Normalizing Ethical Mishaps

[Ethics owners are] enmeshed in organizational cultures that reward metric-oriented and fast-paced work with greater resources, thereby ratcheting up the ability to fit in and ratcheting down the capacity to object. In a situation like that — just as it is with current critical data studies scholarship — it is necessarily challenging to distinguish between success and failure. Thus, the work of ethics consists of working through, and not resolving, these tensions.

Owning ethics: Pitfalls

1. Normalizing Ethical Mishaps

- "fail fast, fail often."

[Ethics owners are] enmeshed in organizational cultures that reward metric-oriented and fast-paced work with greater resources, thereby ratcheting up the ability to fit in and ratcheting down the capacity to object. In a situation like that — just as it is with current critical data studies scholarship — it is necessarily challenging to distinguish between success and failure. Thus, the work of ethics consists of working through, and not resolving, these tensions.

Owning ethics: Pitfalls

1. Normalizing Ethical Mishaps

- "fail fast, fail often."
- "ethics . . . as something to implement rather than something to design organizations around"

[Ethics owners are] enmeshed in organizational cultures that reward metric-oriented and fast-paced work with greater resources, thereby ratcheting up the ability to fit in and ratcheting down the capacity to object. In a situation like that — just as it is with current critical data studies scholarship — it is necessarily challenging to distinguish between success and failure. Thus, the work of ethics consists of working through, and not resolving, these tensions.

Owning ethics: Pitfalls

1. Normalizing Ethical Mishaps

- "fail fast, fail often."
- "ethics . . . as something to implement rather than something to design organizations around"

2. Blinkered Isomorphism

[Ethics owners are] enmeshed in organizational cultures that reward metric-oriented and fast-paced work with greater resources, thereby ratcheting up the ability to fit in and ratcheting down the capacity to object. In a situation like that — just as it is with current critical data studies scholarship — it is necessarily challenging to distinguish between success and failure. Thus, the work of ethics consists of working through, and not resolving, these tensions.

Owning ethics: Pitfalls

1. Normalizing Ethical Mishaps

- "fail fast, fail often."
- "ethics . . . as something to implement rather than something to design organizations around"

2. Blinkered Isomorphism

- "innovation through imitation"

[Ethics owners are] enmeshed in organizational cultures that reward metric-oriented and fast-paced work with greater resources, thereby ratcheting up the ability to fit in and ratcheting down the capacity to object. In a situation like that — just as it is with current critical data studies scholarship — it is necessarily challenging to distinguish between success and failure. Thus, the work of ethics consists of working through, and not resolving, these tensions.

Owning ethics: Pitfalls

1. Normalizing Ethical Mishaps

- "fail fast, fail often."
- "ethics . . . as something to implement rather than something to design organizations around"

2. Blinkered Isomorphism

- "innovation through imitation"
- "perform ethics rather than make the structural changes"

[Ethics owners are] enmeshed in organizational cultures that reward metric-oriented and fast-paced work with greater resources, thereby ratcheting up the ability to fit in and ratcheting down the capacity to object. In a situation like that — just as it is with current critical data studies scholarship — it is necessarily challenging to distinguish between success and failure. Thus, the work of ethics consists of working through, and not resolving, these tensions.

Related: Ethics Theater

> *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

see ethics washing

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).

see ethics washing

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).

see ethics washing

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).

see ethics washing

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).

see ethics washing

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).

see ethics washing

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).

Apex Predator Problem

> *Currently the loudest voices debating the potential dangers of super-intelligence are affluent white men, and, perhaps for them, the biggest threat is the rise of an artificially intelligent apex predator.*

https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html

power and principles

how did new capabilities rearrange power? who can now do what, from what, to whom?

- re-identification

how did new capabilities rearrange power? who can now do what, from what, to whom?

- re-identification
- privacy vs.:

how did new capabilities rearrange power? who can now do what, from what, to whom?

- re-identification
- privacy vs.:
  - face recognition...

how did new capabilities rearrange power? who can now do what, from what, to whom?

- re-identification
- privacy vs.:
  - face recognition...
  - credit scores...

how did new capabilities rearrange power? who can now do what, from what, to whom?

- re-identification
- privacy vs.:
  - face recognition...
  - credit scores...
  - predictive policing from police data...

how did new capabilities rearrange power? who can now do what, from what, to whom?

- re-identification
- privacy vs.:
  - face recognition...
  - credit scores...
  - predictive policing from police data...
  - hiring from hiring data...

how did new capabilities rearrange power? who can now do what, from what, to whom?

- re-identification
- privacy vs.:
  - face recognition...
  - credit scores...
  - predictive policing from police data...
  - hiring from hiring data...
  - contact tracing...

role of rights, harms, justice?

foreshadowing data for Thursday

foreshadowing data for Thursday

- privacy & FAT lab

reminder of themes/big main takeaways

reminder of themes/big main takeaways

- reminder: last week: ML ("AI2.0"), and its impact

reminder of themes/big main takeaways

- reminder: last week: ML ("AI2.0"), and its impact
  - the tech

reminder of themes/big main takeaways

- reminder: last week: ML ("AI2.0"), and its impact
  - the tech
  - interpretability/fairness

reminder of themes/big main takeaways

- reminder: last week: ML ("AI2.0"), and its impact
  - the tech
  - interpretability/fairness
  - role of ethics/privacy

reminder of themes/big main takeaways

- reminder: last week: ML ("AI2.0"), and its impact
  - the tech
  - interpretability/fairness
  - role of ethics/privacy
- today: how researchers *define* ethics (1974-1979)

reminder of themes/big main takeaways

- reminder: last week: ML ("AI2.0"), and its impact
  - the tech
  - interpretability/fairness
  - role of ethics/privacy
- today: how researchers *define* ethics (1974-1979)
  - areas missed by the pre-digital, 1970s approach

reminder of themes/big main takeaways

- reminder: last week: ML ("AI2.0"), and its impact
  - the tech
  - interpretability/fairness
  - role of ethics/privacy
- today: how researchers *define* ethics (1974-1979)
  - areas missed by the pre-digital, 1970s approach
- today: how researchers *design for* ethics

reminder of themes/big main takeaways

- reminder: last week: ML ("AI2.0"), and its impact
  - the tech
  - interpretability/fairness
  - role of ethics/privacy
- today: how researchers *define* ethics (1974-1979)
  - areas missed by the pre-digital, 1970s approach
- today: how researchers *design for* ethics
  - remaining 2 weeks: political dynamics of data ethics IRL

up next

- 4/23: the business of data

up next

- 4/23: the business of data
  - the attention economy, 1971->present

up next

- 4/23: the business of data
  - the attention economy, 1971->present
  - venture: accelerant and destabilizer

up next

- 4/23: the business of data
  - the attention economy, 1971->present
  - venture: accelerant and destabilizer
  - attention economy+venture=information platforms

up next

- 4/23: the business of data
  - the attention economy, 1971->present
  - venture: accelerant and destabilizer
  - attention economy+venture=information platforms
- 4/30: our future

up next

- 4/23: the business of data
  - the attention economy, 1971->present
  - venture: accelerant and destabilizer
  - attention economy+venture=information platforms
- 4/30: our future
  - corporate power

up next

- 4/23: the business of data
  - the attention economy, 1971->present
  - venture: accelerant and destabilizer
  - attention economy+venture=information platforms
- 4/30: our future
  - corporate power
  - state power

up next

- 4/23: the business of data
  - the attention economy, 1971->present
  - venture: accelerant and destabilizer
  - attention economy+venture=information platforms
- 4/30: our future
  - corporate power
  - state power
  - people power

appendix

appendix

- 2020-01-21 : 1 of 14 intro to course

appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes

appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics

appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism

appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy

appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism

appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation

appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation
- 2020-03-10 : 8 of 14 birth and death of AI

appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation
- 2020-03-10 : 8 of 14 birth and death of AI
- 2020-03-24 : 9 of 14 big data, old school (1958-1980)

appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation
- 2020-03-10 : 8 of 14 birth and death of AI
- 2020-03-24 : 9 of 14 big data, old school (1958-1980)
- 2020-03-31 : 10 of 14 data science, 1962-2017

appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation
- 2020-03-10 : 8 of 14 birth and death of AI
- 2020-03-24 : 9 of 14 big data, old school (1958-1980)
- 2020-03-31 : 10 of 14 data science, 1962-2017
- 2020-04-07 : 11 of 14 AI2.0

appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation
- 2020-03-10 : 8 of 14 birth and death of AI
- 2020-03-24 : 9 of 14 big data, old school (1958-1980)
- 2020-03-31 : 10 of 14 data science, 1962-2017
- 2020-04-07 : 11 of 14 AI2.0
- 2020-04-14 : 12 of 14 ethics

appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation
- 2020-03-10 : 8 of 14 birth and death of AI
- 2020-03-24 : 9 of 14 big data, old school (1958-1980)
- 2020-03-31 : 10 of 14 data science, 1962-2017
- 2020-04-07 : 11 of 14 AI2.0
- 2020-04-14 : 12 of 14 ethics
- 2020-04-21 : 13 of 14 present problems: attention economy+VC=dumpsterfire

appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation
- 2020-03-10 : 8 of 14 birth and death of AI
- 2020-03-24 : 9 of 14 big data, old school (1958-1980)
- 2020-03-31 : 10 of 14 data science, 1962-2017
- 2020-04-07 : 11 of 14 AI2.0
- 2020-04-14 : 12 of 14 ethics
- 2020-04-21 : 13 of 14 present problems: attention economy+VC=dumpsterfire
- 2020-04-28 : 14 of 14 future solutions