

data-ppf.github.io 2021-01-19

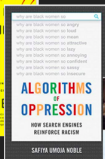
lecture 2 of 14: setting the stakes

chris wiggins + matt jones, Columbia

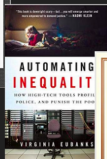
recall from lecture 1



2017-09-05: cathy o'neil



2018-01-08: safiya noble



2018-01-23: virginia eubanks



2019-01-15: shoshana zuboff



2019-06-17: ruha benjamin

something is wrong on the internet

themes for today

themes for today

- ▶ “big data” (the (prescient) view from 2012-2014)

themes for today

- ▶ “big data” (the (prescient) view from 2012-2014)
 - ▶ what are the concerns

themes for today

- ▶ “big data” (the (prescient) view from 2012-2014)
 - ▶ what are the concerns
 - ▶ ethical considerations

themes for today

- ▶ “big data” (the (prescient) view from 2012-2014)
 - ▶ what are the concerns
 - ▶ ethical considerations
 - ▶ political considerations

themes for today

- ▶ “big data” (the (prescient) view from 2012-2014)
 - ▶ what are the concerns
 - ▶ ethical considerations
 - ▶ political considerations
- ▶ “persusasion architectures”

themes for today

- ▶ “big data” (the (prescient) view from 2012-2014)
 - ▶ what are the concerns
 - ▶ ethical considerations
 - ▶ political considerations
- ▶ “persusasion architectures”
 - ▶ considerations for politics

themes for today

- ▶ “big data” (the (prescient) view from 2012-2014)
 - ▶ what are the concerns
 - ▶ ethical considerations
 - ▶ political considerations
- ▶ “persusasion architectures”
 - ▶ considerations for politics
 - ▶ past: (Bernays, etc)

themes for today

- ▶ “big data” (the (prescient) view from 2012-2014)
 - ▶ what are the concerns
 - ▶ ethical considerations
 - ▶ political considerations
- ▶ “persusasion architectures”
 - ▶ considerations for politics
 - ▶ past: (Bernays, etc)
 - ▶ present: challenge to democracy

student observations (quantitative)

student observations (quantitative)

29 political

22 companies

21 privacy

21 ethical

17 bias

10 transparency

10 consequences

9 objectivity

student observations (quotes):

the difference between data being public and data being shared in public

These developments also blur the definitions of consent, anonymity and privacy - is publishing a post on a social media an implicit consent of the usage of the post as a data point? Does registering for an account automatically mean the consent of the user's profile being collected and analyzed?

to engage with social communities in any capacity online, it seems like a requirement to give up your right to privacy.

big data: what is it good for?

My humanities frame of mind is readily catering to doomsday predilections after these readings. I'm not sure I understand the public benefits of big data yet. If our social ills are to a flood as society is to a basement, then it seems like Wallach's computational social science, at best, measures the height of the water rather than helps to drain it.

dystopia

there is a sociological phenomenon, ... people do not follow up on statistics. ... [borne] from our need to consistently be clued in. we are left vulnerable to the crippling of our relationships with friends, family, coworkers, and neighbors—who (thanks to our individually curated algorithms) may see different statistics to us altogether. Worse, we are left dissociated from a common reality and increasingly indebted and entrenched in our constructed reality (re-enter: confirmation bias).

readings: boyd+Crawford, Wallach, Tufekci

boyd+Crawford (2012)

boyd+Crawford (2012)

- ▶ writing from MSRBOS (later MSRNYC)

Technology is neither good nor bad; nor is it neutral . . . technology's interaction with the social ecology is such that technical developments frequently have environmental, social, and human consequences that go far beyond the immediate purposes of the technical devices and practices themselves. (Kranzberg 1986, p. 545)

- ▶ Q: how does 'big data' relate to 'technology?'

boyd+Crawford: themes

1. Big Data changes the definition of knowledge

boyd+Crawford: themes

1. Big Data changes the definition of knowledge
2. Claims to objectivity and accuracy are misleading

boyd+Crawford: themes

1. Big Data changes the definition of knowledge
2. Claims to objectivity and accuracy are misleading
3. Bigger data are not always better data

boyd+Crawford: themes

1. Big Data changes the definition of knowledge
2. Claims to objectivity and accuracy are misleading
3. Bigger data are not always better data
4. Taken out of context, Big Data loses its meaning

boyd+Crawford: themes

1. Big Data changes the definition of knowledge
2. Claims to objectivity and accuracy are misleading
3. Bigger data are not always better data
4. Taken out of context, Big Data loses its meaning
5. Just because it is accessible does not make it ethical

boyd+Crawford: themes

1. Big Data changes the definition of knowledge
2. Claims to objectivity and accuracy are misleading
3. Bigger data are not always better data
4. Taken out of context, Big Data loses its meaning
5. Just because it is accessible does not make it ethical
6. Limited access to Big Data creates new digital divides

b+C: NB

- ▶ attention to PII

b+C: why we chose

Wallach (Dec 2014)

Wallach: background

- ▶ NeurIPS 2014 1st **FATML** meeting

Wallach: background

- ▶ NeurIPS 2014 1st **FATML** meeting
- ▶ Wallach is NLP/CS PhD and prof -> MSRNYC

Wallach: themes

- ▶ Defining big data

Wallach: themes

- ▶ Defining big data
- ▶ Concerns: “ethics. . . outside the expertise”

Wallach: themes

- ▶ Defining big data
- ▶ Concerns: “ethics. . . outside the expertise”
 - ▶ “granularity”

Wallach: themes

- ▶ Defining big data
- ▶ Concerns: “ethics. . . outside the expertise”
 - ▶ “granularity”
- ▶ “Computational Social Science” / cross-disciplinary benefits

Wallach: themes

- ▶ Defining big data
- ▶ Concerns: “ethics. . . outside the expertise”
 - ▶ “granularity”
- ▶ “Computational Social Science” / cross-disciplinary benefits
- ▶ biases: “even scientists”

Wallach: why we chose

- ▶ impact of FATML

Wallach: why we chose

- ▶ impact of FATML
 - ▶ fairness (tech field)

Wallach: why we chose

- ▶ impact of FATML
 - ▶ fairness (tech field)
 - ▶ fatACM

Wallach: why we chose

- ▶ impact of FATML
 - ▶ fairness (tech field)
 - ▶ fatACM
- ▶ Q: where do F,A,T sit in rights, harms, justice?

Wallach: why we chose

- ▶ impact of FATML
 - ▶ fairness (tech field)
 - ▶ fatACM
- ▶ Q: where do F,A,T sit in rights, harms, justice?
- ▶ taxonomy

Wallach: why we chose

- ▶ impact of FATML
 - ▶ fairness (tech field)
 - ▶ fatACM
- ▶ Q: where do F,A,T sit in rights,harms,justice?
- ▶ taxonomy
 - ▶ Wallach: prediction, explanation and exploration

Wallach: why we chose

- ▶ impact of FATML
 - ▶ fairness (tech field)
 - ▶ fatACM
- ▶ Q: where do F,A,T sit in rights, harms, justice?
- ▶ taxonomy
 - ▶ Wallach: prediction, explanation and exploration
 - ▶ Gartner: describe, predict, prescribe

Tufekci

Tufekci: themes

► *past*

Tufekci: themes

- ▶ *past*
 - ▶ Bernays

Tufekci: themes

- ▶ *past*
 - ▶ Bernays
 - ▶ 1968 (and earlier)

Tufekci: themes

- ▶ *past*
 - ▶ Bernays
 - ▶ 1968 (and earlier)
- ▶ *future*

Tufekci: themes

- ▶ *past*
 - ▶ Bernays
 - ▶ 1968 (and earlier)
- ▶ *future*
 - ▶ 2016 etc.

Tufekci: past

- ▶ Bernays

Tufekci: past

- ▶ Bernays
 - ▶ [Engineering of Consent](#) 1947/1955 “engineering approach. . . scientific principles. . . getting people to support ideas and programs”

Tufekci: past

- ▶ Bernays
 - ▶ [Engineering of Consent](#) 1947/1955 “engineering approach. . . scientific principles. . . getting people to support ideas and programs”
- ▶ For more info, enjoy

Tufekci: past

- ▶ Bernays
 - ▶ [Engineering of Consent](#) 1947/1955 “engineering approach. . . scientific principles. . . getting people to support ideas and programs”
- ▶ For more info, enjoy
 - ▶ [Century of the Self](#) (2002, 4 hrs)

Tufekci: past

- ▶ Bernays
 - ▶ [Engineering of Consent](#) 1947/1955 “engineering approach. . . scientific principles. . . getting people to support ideas and programs”
- ▶ For more info, enjoy
 - ▶ [Century of the Self](#) (2002, 4 hrs)
 - ▶ [These Truths](#) and “[Politics and the New Machine](#)” from [Jill Lepore](#) (esp. on [Simulmatics](#))

Tufekci: Bernays as backdrop

Intelligent men must realize that propaganda is the modern instrument by which they can fight for productive ends and help to bring order out of chaos

- Bernays
 - ▶ role of experts in science vs morals

Tufekci: Bernays as backdrop

Intelligent men must realize that propaganda is the modern instrument by which they can fight for productive ends and help to bring order out of chaos

- Bernays
 - ▶ role of experts in science vs morals
 - ▶ dark money, 1920s style

Tufekci: more on Bernays

There was one basic lesson I learned in the CPI—that efforts comparable to those applied by the CPI to affect the attitudes of the enemy, of neutrals, and people of this country could be applied with equal facility to peacetime pursuits. In other words, what could be done for a nation at war could be done for organizations and people in a nation at peace.

- Bernays, 1891-1995, quoted in [Cutlip](#), 1994, on his WWI work with the [Committee on Public Information](#)

Tufekci: structure/summary

1. “Big data”: (“especially to large organizations”)

Tufekci: structure/summary

1. “Big data”: (“especially to large organizations”)
2. “methods”: “allow political [individual] targeting”

Tufekci: structure/summary

1. “Big data”: (“especially to large organizations”)
2. “methods”: “allow political [individual] targeting”
3. “modeling”: prediction “and opacity.”

Tufekci: structure/summary

1. “Big data”: (“especially to large organizations”)
2. “methods”: “allow political [individual] targeting”
3. “modeling”: prediction “and opacity.”
4. “behavioral sciences”: “irrational. . . empirical. . . psychographic”

Tufekci: structure/summary

1. “Big data”: (“especially to large organizations”)
2. “methods”: “allow political [individual] targeting”
3. “modeling”: prediction “and opacity.”
4. “behavioral sciences”: “irrational. . . empirical. . . psychographic”
5. “experimental. . . real-time”: “molding the public sphere”

Tufekci: structure/summary

1. “Big data”: (“especially to large organizations”)
2. “methods”: “allow political [individual] targeting”
3. “modeling”: prediction “and opacity.”
4. “behavioral sciences”: “irrational. . . empirical. . . psychographic”
5. “experimental. . . real-time”: “molding the public sphere”
6. “platforms”: “proprietary. . . opaque algorithms”

Tufekci: why we chose

- ▶ stakes:

Tufekci: why we chose

- ▶ stakes:
 - ▶ data and the public esp. political

Tufekci: why we chose

- ▶ stakes:
 - ▶ data and the public esp. political
 - ▶ cf. data and “research” (focus in Wallach, boyd+Crawford)

Tufekci: why we chose

- ▶ stakes:
 - ▶ data and the public esp. political
 - ▶ cf. data and “research” (focus in Wallach, boyd+Crawford)
 - ▶ data and capitalism, race, inequality. . . . (cf. booklist)

Tufekci: why we chose

- ▶ stakes:
 - ▶ data and the public esp. political
 - ▶ cf. data and “research” (focus in Wallach, boyd+Crawford)
 - ▶ data and capitalism, race, inequality. . . . (cf. booklist)
- ▶ primary sources of blame:

Tufekci: why we chose

- ▶ stakes:
 - ▶ data and the public esp. political
 - ▶ cf. data and “research” (focus in Wallach, boyd+Crawford)
 - ▶ data and capitalism, race, inequality. . . . (cf. booklist)
- ▶ primary sources of blame:
 - ▶ researchers (Wallach, boyd+crawford)

Tufekci: why we chose

- ▶ stakes:
 - ▶ data and the public esp. political
 - ▶ cf. data and “research” (focus in Wallach, boyd+Crawford)
 - ▶ data and capitalism, race, inequality. . . . (cf. booklist)
- ▶ primary sources of blame:
 - ▶ researchers (Wallach, boyd+crawford)
 - ▶ platform companies (Tufekci *inter alia*)

Tufekci: why we chose

- ▶ stakes:
 - ▶ data and the public esp. political
 - ▶ cf. data and “research” (focus in Wallach, boyd+Crawford)
 - ▶ data and capitalism, race, inequality. . . . (cf. booklist)
- ▶ primary sources of blame:
 - ▶ researchers (Wallach, boyd+crawford)
 - ▶ platform companies (Tufekci *inter alia*)
 - ▶ public-private partnerships (Eubanks, O’Neil)

Tufekci: why we chose

- ▶ stakes:
 - ▶ data and the public esp. political
 - ▶ cf. data and “research” (focus in Wallach, boyd+Crawford)
 - ▶ data and capitalism, race, inequality. . . . (cf. booklist)
- ▶ primary sources of blame:
 - ▶ researchers (Wallach, boyd+crawford)
 - ▶ platform companies (Tufekci *inter alia*)
 - ▶ public-private partnerships (Eubanks, O’Neil)
- ▶ how did platform companies respond?

Tufekci: why we chose

- ▶ stakes:
 - ▶ data and the public esp. political
 - ▶ cf. data and “research” (focus in Wallach, boyd+Crawford)
 - ▶ data and capitalism, race, inequality. . . . (cf. booklist)
- ▶ primary sources of blame:
 - ▶ researchers (Wallach, boyd+crawford)
 - ▶ platform companies (Tufekci *inter alia*)
 - ▶ public-private partnerships (Eubanks, O’Neil)
- ▶ how did platform companies respond?
 - ▶ A: wait until week 12-14 to find out

Tufekci: why we chose

- ▶ stakes:
 - ▶ data and the public esp. political
 - ▶ cf. data and “research” (focus in Wallach, boyd+Crawford)
 - ▶ data and capitalism, race, inequality. . . . (cf. booklist)
- ▶ primary sources of blame:
 - ▶ researchers (Wallach, boyd+crawford)
 - ▶ platform companies (Tufekci *inter alia*)
 - ▶ public-private partnerships (Eubanks, O’Neil)
- ▶ how did platform companies respond?
 - ▶ A: wait until week 12-14 to find out
- ▶ how “new” is data+social statecraft

Tufekci: why we chose

- ▶ stakes:
 - ▶ data and the public esp. political
 - ▶ cf. data and “research” (focus in Wallach, boyd+Crawford)
 - ▶ data and capitalism, race, inequality. . . . (cf. booklist)
- ▶ primary sources of blame:
 - ▶ researchers (Wallach, boyd+crawford)
 - ▶ platform companies (Tufekci *inter alia*)
 - ▶ public-private partnerships (Eubanks, O’Neil)
- ▶ how did platform companies respond?
 - ▶ A: wait until week 12-14 to find out
- ▶ how “new” is data+social statecraft
 - ▶ A: wait until week 3 to find out

Tufekci: why we chose

- ▶ stakes:
 - ▶ data and the public esp. political
 - ▶ cf. data and “research” (focus in Wallach, boyd+Crawford)
 - ▶ data and capitalism, race, inequality. . . . (cf. booklist)
- ▶ primary sources of blame:
 - ▶ researchers (Wallach, boyd+crawford)
 - ▶ platform companies (Tufekci *inter alia*)
 - ▶ public-private partnerships (Eubanks, O’Neil)
- ▶ how did platform companies respond?
 - ▶ A: wait until week 12-14 to find out
- ▶ how “new” is data+social statecraft
 - ▶ A: wait until week 3 to find out
 - ▶ spoliter: is as old as “social science”, is “older than statistics”

Tufekci: why we chose

- ▶ stakes:
 - ▶ data and the public esp. political
 - ▶ cf. data and “research” (focus in Wallach, boyd+Crawford)
 - ▶ data and capitalism, race, inequality. . . . (cf. booklist)
- ▶ primary sources of blame:
 - ▶ researchers (Wallach, boyd+crawford)
 - ▶ platform companies (Tufekci *inter alia*)
 - ▶ public-private partnerships (Eubanks, O’Neil)
- ▶ how did platform companies respond?
 - ▶ A: wait until week 12-14 to find out
- ▶ how “new” is data+social statecraft
 - ▶ A: wait until week 3 to find out
 - ▶ spoliter: is as old as “social science”, is “older than statistics”
 - ▶ next week: social physics and statistics as quantitative statecraft

power and principles

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

role of rights, harms, justice?

foreshadowing data for Thursday

reminder of themes/big main takeaways

themes

up next

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: 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: data science, 1962-2017
- ▶ 2021-03-30: 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: data science, 1962-2017
- ▶ 2021-03-30: AI2.0
- ▶ 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: data science, 1962-2017
- ▶ 2021-03-30: AI2.0
- ▶ 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: data science, 1962-2017
- ▶ 2021-03-30: AI2.0
- ▶ 2021-04-06: ethics
- ▶ 2021-04-13: present problems: attention
economy+VC=dumpsterfire
- ▶ 2021-04-15: future solutions