GOOD SCIENCE? INTRODUCTION TO ETHICS FOR DATA SCIENCE

PROF. CHARIS THOMPSON

CHANCELLOR'S PROFESSOR

UC BERKELEY AND LSE

APRIL 20TH, 2017, 306 SODA,12:30-2:00

ETHICS APPROACHES: WHERE DO THEY LOCATE THE GOOD, AND DO THEY WORK FOR DATA SCIENCE?

- **Utilitarian / consequentialist ethics** (what's the best outcome in terms of e.g. happiness, pleasure, pain, life and flourishing?) "The end justifies the means"

- **Deontological / principle-based / rule following ethics** (is it right, based on a fundamental principle? did you do your duty, and follow the rules?) "This is a good in and of itself"; "If you follow the rules, you can't go wrong"

- **Virtue ethics** (are you / your character / your mind good?) "Are you a good person?"

- **Pragmatic ethics** (open to empirical, scientific inquiry and can be changed over time and place if it turns out to be wrong) "It turns out that X is good for Y"

- **Local moral worlds** (what are acceptable norms where we live today, how do we respect them, and when do we change them?) "Here in Berkeley, we should take into consideration X, Y, Z, because of our history and because these things matter to us"
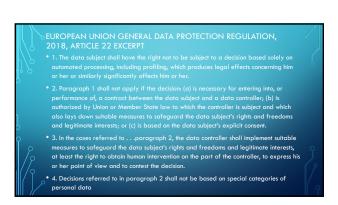

WHERE ELSE COULD WE LOOK? E.G. SHOULD DATA SCIENCE ETHICS DRAW ON US MEDICAL ETHICS?

- In the US, in the postwar period, building on the Nuremberg Code (1947) and other ethical instruments, four principles of bioethics have stabilized:

- **Respect for autonomy**: That people be informed (informed consent) and that they make decisions without coercion and can opt out of research or experiments, which has a kind of echo in right to privacy and right to being forgotten

- **Non-maleficence**: the idea in medical ethics of "first do no harm". For big data, we are in a phase where certain basic parameters are being laid down

- **Beneficence**: that you aim to do actual good, and that good outweighs harm

- **Justice**: fairness in how you are treated and that you are neither more nor less likely to be a subject of research than anyone else based on group memberships or personal characteristics

---

ETHICAL CONCERNS FOR DATA SCIENCE:

- **Data ownership** – at source and up and down the value chain (e.g. genomic information)
- **Informed Consent** – who has to get it, who has to give it, and why (e.g. We Vibe 4 Plus vibrator)
- **Privacy** – what is it and when and how should it be granted (e.g. EURODAC, IAFIS biometric databases)
- **Fairness** – making sure data neither over nor under represents anyone, and does not discriminate against some groups (e.g. redlining)
- **Cybersecurity** – making sure that data are secure; risks of not doing do can be catastrophic (e.g. e.g. finance, national security, public health)
- **Democracy** – enhancing rather than reducing literacy and numeracy and engagement (e.g. bubble / echo chambers)
- **Changes to the economy** – as machine learning takes over, how are our lives being fundamentally restructured (e.g. the future of work and recalcitrant inequality)

FAIRNESS AS SOCIAL JUSTICE: RACISM, SEXISM, CLASSISM, ABLEISM, URBAN BIAS, FIRST WORLD BIAS

- Dave Coplin, Microsoft: "in AI every time an algorithm is written, embedded within it will be all the biases that exist in the humans who created it". We need "to be mindful of the philosophies, morals and ethics of the organizations […] creating the algorithms that increasingly we rely on every day" (House of Commons report, 10/16)

- Discrimination can result from either completely "neutral" things like "uncertainty bias" where underrepresented groups are riskier because based on fewer data points, or embedded bias, such as word association vectors

- The **good news** is that algorithms are not invested in bias perpetuation, but the **bad news** is black-boxing and the challenge of e.g. natural language learning any other way than on data sets that contain our biases

- Many scholars, many terms for this, e.g.: Safiya Umoja Noble **algorithms of oppression**; Cathy O'Neil **weapons of math destruction**; Holston, Ochigame **algorithmic filtering**

EUROPEAN UNION GENERAL DATA PROTECTION REGULATION, 2018, ARTICLE 22 EXCERPT

- 1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

- 2. Paragraph 1 shall not apply if the decision: (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller; (b) is authorized by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or (c) is based on the data subject's explicit consent.

- 3. In the cases referred to . . .paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

- 4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data

LOCAL, NATIONAL AND GLOBAL GOVERNANCE CONCERNS

- Data is used to open and close borders; it also crosses borders, opening up not only security and human rights risks but also massive governance challenges

- Coming up with guidelines and regulation for different jurisdictions is tough enough

- Harmonizing regulations across jurisdictions is extraordinarily difficult

- Compliance is arguably harder still – what are the compliance mechanisms in each jurisdiction and how are multi-jurisdictional mechanisms to be funded, coordinated, and so on

- 4th Industrial Revolution concerns – as life sciences and digital and material sciences are merging, how do we track the new world?

- Militarism concerns – the era of "everything is dual use" from autonomous weapons to civilian drones

SCIENCE FARE: SOCIAL BENCHMARKING AND PUBLIC ENGAGEMENT

- Building in social goals in formal specifications; measure and correct shortfalls

- Infrastructure and goals: social benchmarks and milestones

- Empaneling experts and non-experts, stakeholders and non-stakeholders to set goals and measure outcomes and set correctives

- Those historically underserved within different arenas, such as underrepresented minorities and healthcare or financial data; disability justice scholars and activists setting goals and correctives and monitoring for assistive and augmentative devices

- Highlight projects and set goals at the interface of social justice and data: e.g. building up 360 and panoramic views from e.g. police body cameras