Ethics in Data Science
Metis, SF

Nov. 9, 2017

Ovid Jacob

Introduction

The (Big) Data revolution raises a number of ethical issues related to privacy, confidentiality, transparency and identity.

Who owns all that data that we are analyzing? Are there limits to what kinds of inferences we can make, or what decisions can be made about people based on those inferences?

People have too much trust in numbers to be intrinsically objective - Cathy O'Neill

What is Ethics?

As a branch of philosophy, ethics investigates the questions "What is the best way for people to live?" and "What actions are right or wrong in particular circumstances?" In practice, ethics seeks to resolve questions of human morality, by defining concepts such as good and evil, right and wrong, virtue and vice, justice and crime.

Discussion: other definitions of Ethics

Who is responsible for how data is used?

Managers at social media?

Engineers at social media?

Users?

Others?

Discussion on these agents

Facets of ethics in data science

Social

Psychological/Psychometrical

Economic

Technical

Political

Legal

Medical

"Amid the panic and confusion surrounding at hostages-taking in Sydney, Australia, Uber - the hipster taxi service that insists that it is not a taxi service - did the unconscionable: it raised the price of getting a lift out of the danger zone" - LA Times

"But it didn't just raise the price. Uber's automatic "surge pricing" model gouged the panic-stricken to the tune of more than $100 ride, a quadrupling of the regular rate." - LA Times

Discuss ethics of surge pricing. Is it ever OK?

Rachel Maddow on MSNBC on Social Media companies and flaws in ML/AI, Oct.10, 2017

https://youtu.be/gGLLL_Wqa4M

This is a story of social media co.'s seeming collusion w Russian propaganda.
Or great reluctance to come out w the facts, even now. Possible reasons:

1. loathing to admit incompetence

2. web traffic drives revenue. So from that point of view. no 'bad' web traffic

3. not wanting to open possibility of 'nationalizing' social media, of regulation

Recall MS's AI bot that was highjacked by hateful trolls (March 2016):

http://www.techrepublic.com/article/why-microsofts-tay-ai-bot-went-wrong/

Maybe social media has likewise been highjacked by hateful trolls.

Is social media 'sticky' by design? Discussion

Cambridge Analytica

Uses of psychometrics in politics

Examples: Brexit, Ted Cruz's campaign, Donald J Trump's campaign

Discuss the ethics of narrowcasting advertising/propaganda

Several aspects of ethics in data science:

Data acquisition

Data analysis

Algorithms

Ethical Issues in Data Acquisition

What kind of data is there?

Discuss types of data: personal private, personal public, device generated, social, meta-data from internet usage, Internet of Things

How is data acquired?

Voluntary, freely given

Involuntary, given unknowingly.

Discuss data acquisition.

Separate collection of personal data from collection of device data. Maybe discard collection of personal data altogether.

Discuss strategies for maintaining privacy in data collection in view of the Equifax data breach

3rd party doctrine

3rd party doctrine makes personal data public, since users have agreed to shared it w 3rd party ISPs. It is no longer private, therefore available for use by gov't agencies.

Discussion: Has this issue really been decided?

Who owns the data we create in social media? We? The social media?

Should we be paid for giving our data to the social platform we use? How much?

Discuss who owns the data we create and how to estimate its value

What is the value of our data?

Per user value?

Collective value?

Discussion

How can changes be made to data policy if the social media is private?

Discussion: regulations/laws governing social media?

Discussion: should social media stay private, or be regulated, like a utility?

Discussion: media publicly owned?

Recent articles on size and dominance of social media by Farhad Manjoo in New York Times

His series of articles on the "Frightful Five" - Amazon, Apple, Facebook, Google, Microsoft, from Jan. 2016 until now. Book to be published soon.

Their size and dominance of these five companies raises the question of whether they should remain private or should there be a public oversight

Discussion on public ownership/control/regulation of social media.

Perhaps we need to revisit our antitrust laws and revive the original focus on the process of unjust wealth redistribution that characterized the Sherman Antitrust act of 1890.

See the work of Lina Khan in the Yale Law Review on this in Jan., 2017 on Amazon's Antitrust Paradox:

Focus on price has missed the amassing of influence Amazon has been able to get, while avoiding antitrust scrutiny. We should replace the consumer welfare framework with an approach oriented around preserving a competitive process and market structure.

https://www.yalelawjournal.org/note/amazons-antitrust-paradox

Privacy of personal data

Do we value privacy of our data?

How do we maintain our privacy online?

Discuss privacy of personal data

Ethical Issues in Data Analysis

Are there biases in data analysis?

What about the issue of prior assumptions in Bayesian probability?

Discussion

Ethical Issues In Algorithms Used in Data Science

There is the issue of false positives, from incomplete correlations. Just because a person does searches on murdering a person, and such items, it does not mean that the person is planning a murder. S/he might be a crime writer!

The idea of pre-crime, like in the movie Minority Report, can lead to many false positives.

Is an algorithm ever wrong?

What about the Trending algorithm on Twitter?

There was some discussion around #occupywallstreet and possible censorship by Twitter of those tweets

What about gender bias or racial bias in algorithms?

Discuss other examples of social bias in algorithms

Response(s)

Privacy-aware design (Prof. Stephen Wicker, Cornell)Ethically-aware data science: in data collection, in data analysis

Privacy-aware data design:

Provide full data collection disclosure

Require consent

Minimize collection

Minimize identification

Minimize and secure retained data

Discussion: other responses

Zero knowledge proof systems

Could provide alternative to present systems.

Discuss other alternatives