

Large Synoptic Survey Telescope Data Science Fellowship Program

5 key questions for human-centered data science

Laura Norén
laura.noren@nyu.edu
















4 May 2018



NEW YORK UNIVERSITY

Moral ranking of occupations (US)

[Gallup Poll "Honesty/Ethics in professions"](#)

Occupation	Gallup Polling, 1985-2015		1985	2015	DELTA
CLERGY			67	45	
PHARMACISTS			65	68	
MEDICAL DOCTORS			58	67	
DENTISTS			56	62	
COLLEGE TEACHERS			54	53	
ENGINEERS			53	70	
POLICE			47	56	
BANKERS			37	25	
TV REPORTERS/COMMENTATORS			33	20	
FUNERAL DIRECTORS			32	44	
JOURNALISTS			31	27	
NEWSPAPER REPORTERS			29	21	
LAWYERS			27	21	
EXECUTIVES			23	17	
MEMBERS OF CONGRESS			20	8	
CAR SALESPeOPLE			5	8	

Pattern detected?
People who talk for
a living are
perceived to be less
honest/ethical.

Moral ranking of occupations (US)

Gallup Poll "Honesty/Ethics in
professions"

Occupation	Gallup Polling, 1985-2015		1985	2015	DELTA
CLERGY			67	45	
PHARMACISTS			65	68	
MEDICAL DOCTORS			58	67	
DENTISTS			56	62	
COLLEGE TEACHERS			54	53	
ENGINEERS			53	70	
POLICE			47	56	
BANKERS			37	25	
TV REPORTERS/COMMENTATORS			33	20	
FUNERAL DIRECTORS			32	44	
JOURNALISTS			31	27	
NEWSPAPER REPORTERS			29	21	
LAWYERS			27	21	
EXECUTIVES			23	17	
MEMBERS OF CONGRESS			20	8	
CAR SALESPeOPLE			5	8	

Does data science need special ethical consideration?

Mostly, the answer is "no".



Ethical training is not new

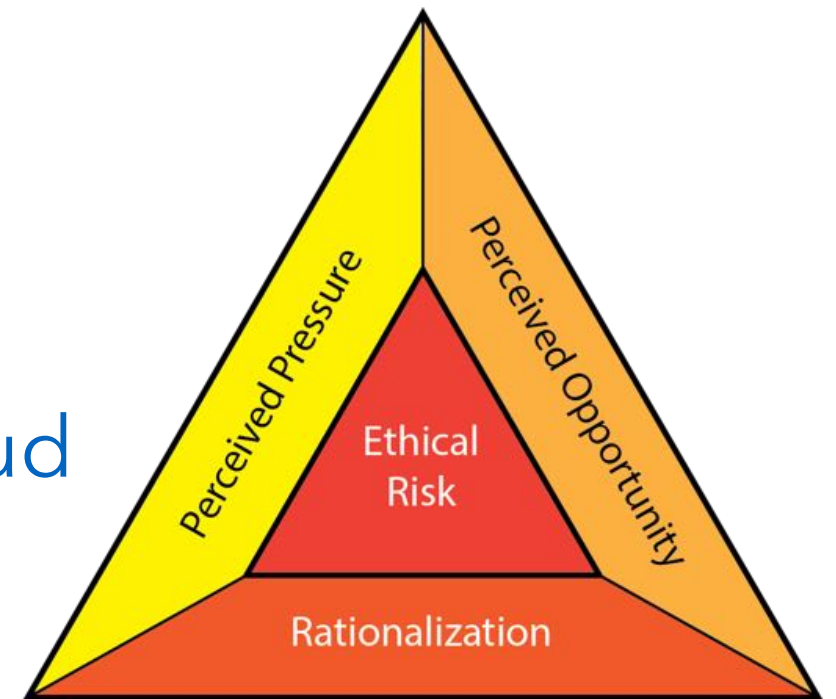
Engineering and business schools already offer ethical training. It covers...

Structural conditions leading to fraud

When/how to blow the whistle

Choose projects wisely: do well (\$) by doing good

Balancing tech/science advance against social harms



But data science *is* new and different...

Data can approach complete representations of humans.

Data persists, it doesn't decay on social time scales.

People don't share tasks well with automated processes.

Algorithmic decision making is often opaque, inexplicable.

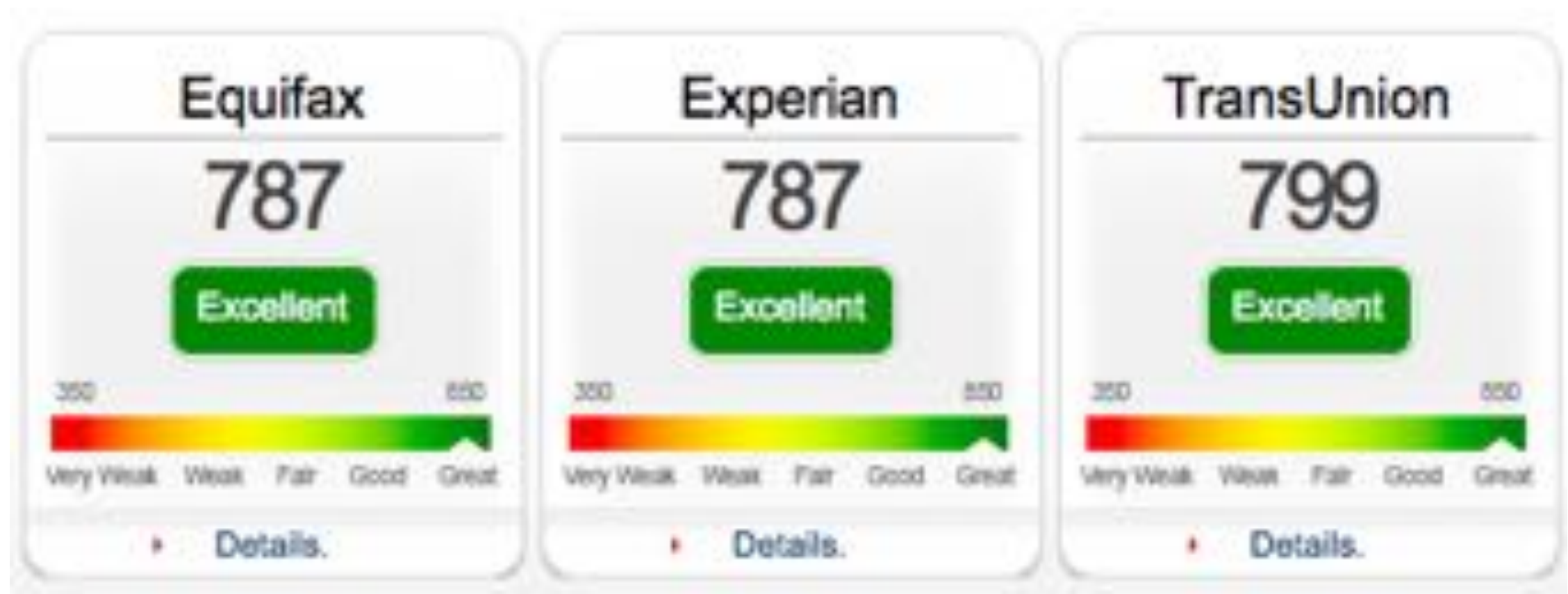
Training data generated by humans will contain bias.

Automated processes will replicate, and occasionally amplify bias.

Bad actors have new, augmented capacity to wreak havoc.

From privacy to power dynamics

I argue that a fruitful way to address these issues is by setting the privacy framework aside and assessing power dynamics.



What people worry about when they worry about privacy

1. They don't know what decisions are made about them or for which purposes.
2. They can't choose which areas of their lives they want to share with companies and automated processes.
3. They have little meaningful choice in data sharing relationships.

Industries and countries are addressing these issues

The EU General Data Protection Regulation emphasizes user consent, gives users some rights to have their data deleted.

Non-compliant companies are fined €20m or 4% of revenue.



Concern: neo-colonialism

Morality and ethical expectations vary dramatically from country to country.

Global multi-national companies are navigating culturally fraught terrain.

Five ethics questions for data science

1. How do we scale consenting procedures?
2. Should data be designed to decay?
3. Should we prevent intelligence inequality?
4. Can we build fail-tolerant machine + human assemblages?
5. Are data science processes further amplifying disadvantage for the poor and otherwise disadvantaged groups?

Scaling consent

We don't have great models for meaningfully consenting people at scale, over time.

Terms of service and privacy policies don't work, nobody reads them.

We need practices that work at scale and across time.

Two proposals for consenting

Ongoing consenting relationships:

Obtain consent little by little for specific kinds of data over time.

Make companies agree to users' terms of service:

Prepare simple to understand, legally valid gradations of consent that users choose to manage their relationship with companies.

Decaying data

Data will last longer than human memory, longer than human life and approach a completeness of representation that is new.

Are there contexts in which data should be designed to decay?

Could we replace a right to be forgotten with an automated habit of selective memory?

Intelligence inequality

The new digital divide will be when neural implants allow some of us to enhance our cognitive capacity.

How do we protect people with neural implants from attack?

How do prevent deeply insidious divide between cognitive haves and have nots?

Making fail tolerant human/machine assemblages

What design principles can we follow to ensure that humans don't:

- take all the blame - becoming moral crumple zones for human/machine system failure (Elish)
- cause or suffer from harms when they share responsibilities with automated systems (Bainbridge)

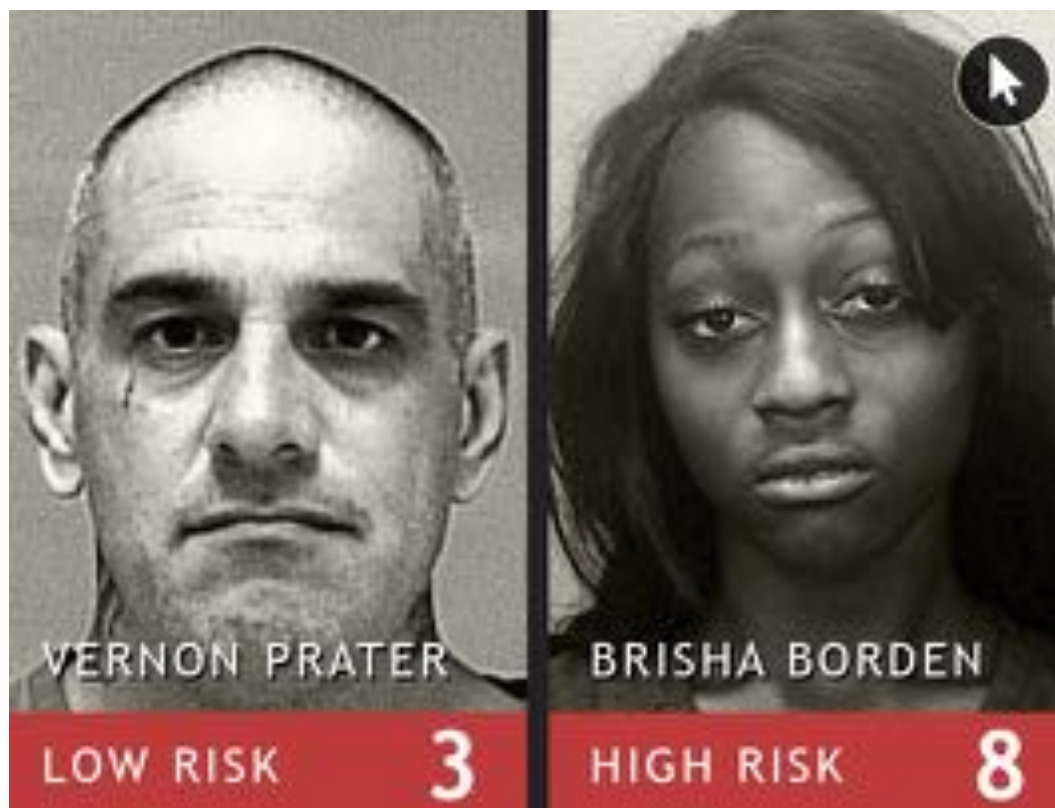
Making fail tolerant human/machine assemblages



Large Synoptic Survey Telescope Corp. - Data Science Fellowship Program

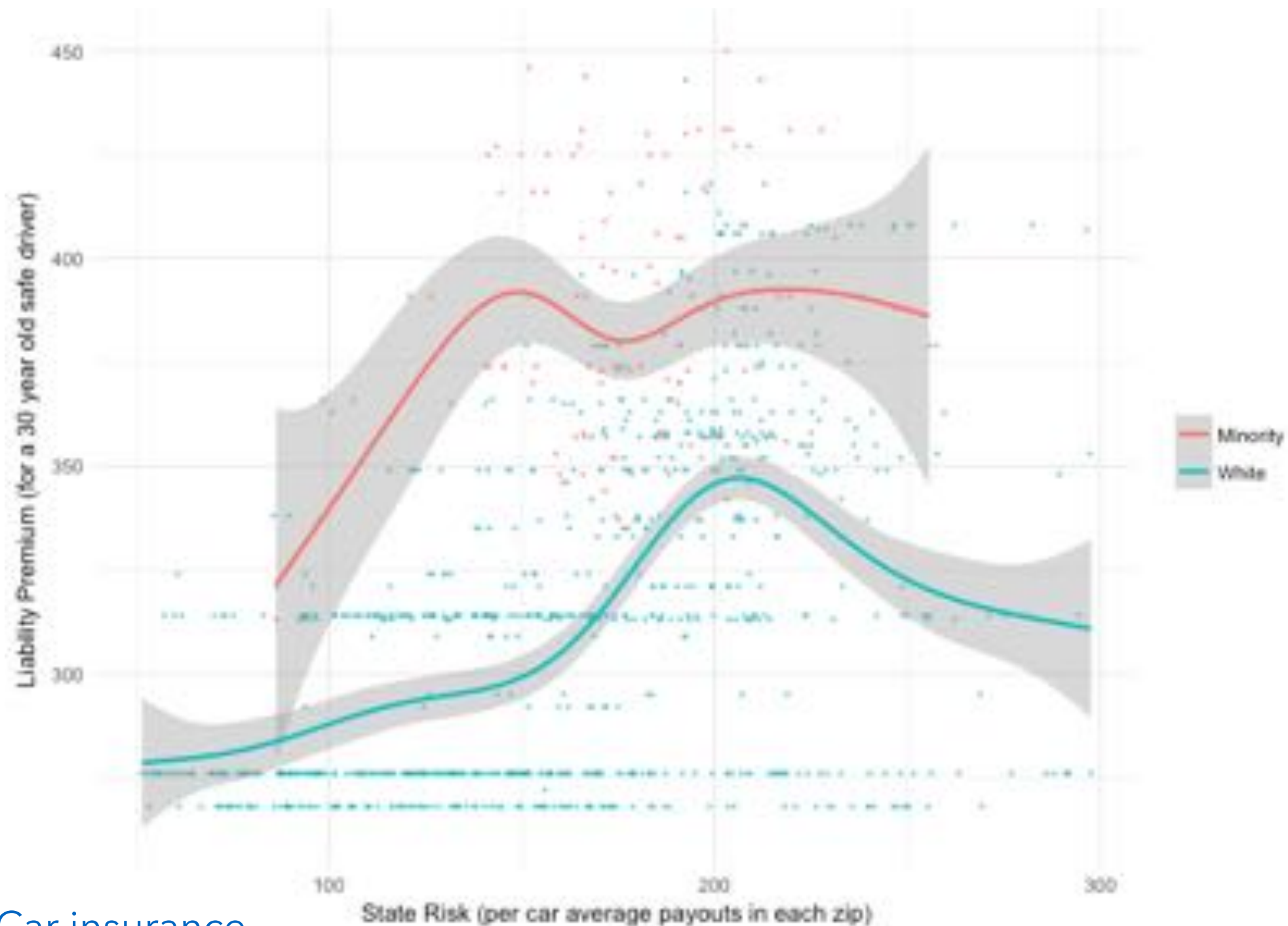
Protect disadvantaged classes

Algorithms that rely on training data will replicate, if not amplify, human biases baked into the initial data set.



Bail sentencing

Protect disadvantaged classes



Car insurance

Large Synoptic Survey Telescope Corp. - Data Science Fellowship Program

Conclusion: Just questions

The field of data science ethics is young.

Researchers are working on formulating questions, providing case studies, and developing tools to avoid certain kinds of bias.

Policy makers are producing strategies for assessing algorithmic bias, developing policy, complying with new policy (GDPR), and establishing auditing rules.

Break

Facebook



Is Facebook responsible for the content users encounter on its site? What about the news?

“individuals don’t control everything they see on Facebook — Facebook’s algorithms do. Facebook is the only one with the power to control what's on its site. And for a company that crows over its abilities to convince people to buy products, play games and even convince people to vote, it rings false for Facebook to reject the idea that its algorithms pick couldn't possibly influence those votes.”

-Hayley Tsukuyama, *The Washington Post*



Mark Zuckerberg

“After the election, many people are asking whether fake news contributed to the result, and what our responsibility is to prevent fake news from spreading. These are very important questions and I care deeply about getting them right.

“Of all the content on Facebook, more than 99% of what people see is authentic. Only a very small amount is fake news and hoaxes. [...]

“In my experience, people are good, and even if you may not feel that way today, believing in people leads to better results over the long term.”

-Mark Zuckerberg, *Facebook post November 2016*

Pizzagate



I won't stop tweeting about [#PizzaGate](#) until I know for a fact that there aren't children in danger being covered up by the US government

— Skye (@TradSierraHotel) Nov. 20, 2016



Facebook responds with a tool



Disputed by Multiple, Independent Fact-Checkers

Before you share this content, you might want to know that the fact-checking sites, Snopes.com and Associated Press disputed its accuracy.

About Disputed Content

Cancel

Post Anyway



Facebook responds with a tool

 Create a Post |  Photo/Video Album 



<http://www.thenewportbuzz.com/the-irish-slave-trade-the-slaves-that-time-forgot/7191>



The Irish Slave Trade – The Slaves That T... 

by G John Sit We've all been taught the horror's of the African...

THENEWPORTBUZZ.COM

 **Disputed by Snopes.com and Associated Press**

Angela Merkel



“Angela Merkel's [German] government has just approved plans which could see social networks fined up to 50m euros if they fail to remove illegal content within 24 hours.”

-BBC



Enter: Cambridge Analytica

Cambridge Analytica is a consulting firm that works on political campaigns, mostly conservative.







Pre-Cambridgian era

Professor Aleksandr Kogan* publishes with
Facebook, Harvard researchers

On wealth and the diversity of friendships: High
social class people around the world have fewer
international friends

Maurice H. Yearwood ^a  , Amy Cuddy ^b, Nishtha Lamba ^a, Wu Youyou ^a, Ilmo van
der Lowe ^a, Paul K. Piff ^c, Charles Gronin ^d, Pete Fleming ^d, Emiliana Simon-Thomas ^c,
Dacher Keltner ^c, Aleksandr Spectre ^{a, 1}

 **Show more**

<https://doi.org/10.1016/j.paid.2015.07.040>

[Get rights and content](#)

Open Access funded by Economic and Social Research Council

Under a Creative Commons [license](#)

[open access](#)



* before getting married and changing his name to Aleksander Spectre

Cambridge Analytica



Professor Aleksander Kogan prepares psychological profile “quizzes” on Facebook’s platform in 2014.

- they build psychometric profiles

- they capture users social graphs

- they capture data from friends’ social graphs



Cambridge Analytica

Professor Kogan prepares the quizzes as part of his research and part of consulting work with Cambridge Analytica.

From 207,000 'consented' users taking the quiz, Cambridge Analytica builds a dataset of 87 million users (social graphs of ~420 people per consented user).

Cambridge Analytica



Kogan's language in the informed consent for the app he built states that he will be using the data for academic research.

The app contains University of Cambridge logo.



UNIVERSITY OF
CAMBRIDGE



Cambridge Analytica

Cambridge Analytica works on behalf of Robert and Rebekah Mercer, politically conservative billionaires, who wish to influence elections.



Facebook



Tech reporters talk to whistleblower and reveal the relationship between Cambridge Analytica, Facebook, and the Mercers.

How Trump Consultants Exploited the Facebook Data of Millions

[Leer en español](#)

By MATTHEW ROSENBERG, NICHOLAS CONFESSORE and CAROLE CADWALLADR MARCH 17, 2018



Facebook



Consumers get mad at Facebook for a “data breach”. Facebook insists it was not a data breach. A data ‘breach’ would have had to be disclosed to users, investors.



Facebook



"This was a major breach. I'm really sorry this happened."

-Zuckerberg

21 Mar 2018



Zuckerberg



Zuck testifies in front of Congress. Many in Congress do not seem to understand how social media or data science work.

House committee will summon Zuckerberg to testify

By JOHN HENDEL | 03/22/2018 01:00 PM EDT



Share on Facebook



Share on Twitter

The House Energy and Commerce Committee will summon Facebook CEO Mark Zuckerberg to testify, following recent revelations that Trump-linked Cambridge Analytica improperly obtained information on some 50 million Facebook users.



Issues at stake:

In the news:

Kogan lied

CA is bad

Facebook had a data breach

Ethically:

Informed consent is broken

We have no clear framework for the ways platforms
with users generating data for sale/use by third parties
ought to operate



Summary of new ethical territory:

The case challenges traditional business ethics that assumes most corporate relationships are dyadic - companies make products used by their customers.

COMPANY ● — ● CUSTOMER

The Facebook - Cambridge Analytica case demonstrates the difficulties associated with the triadic nature of data driven, ad-fueled platforms:

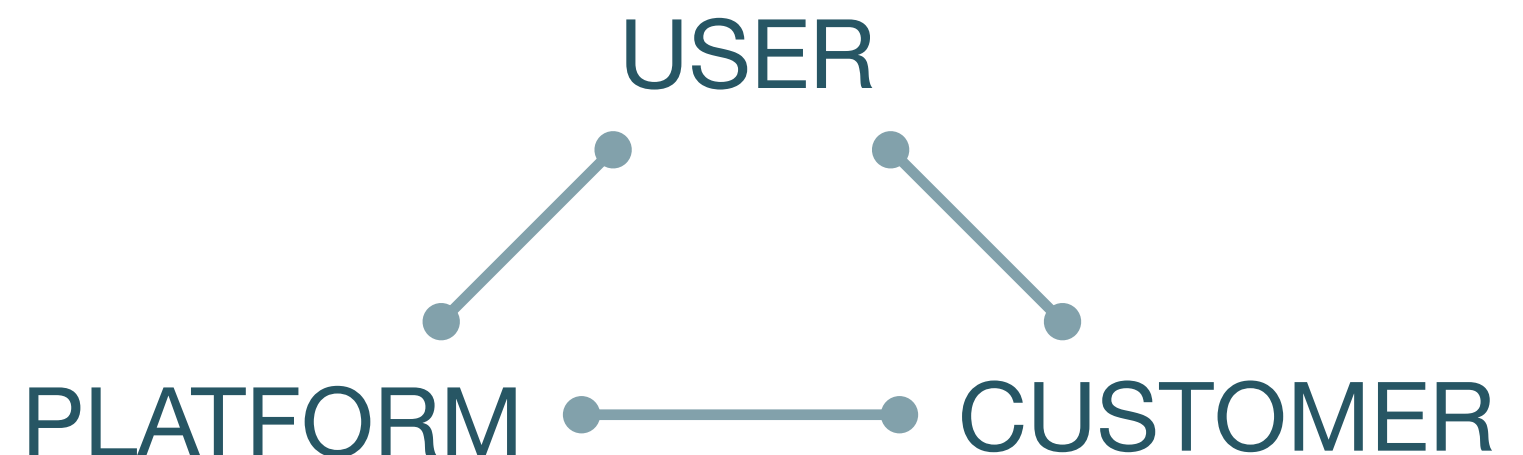
USER ● — ● PLATFORM ● — ● CUSTOMER



Main ethical question:

How do we balance the rights of users in relationships with platforms and companies?

USER ●————● PLATFORM ●————● CUSTOMER





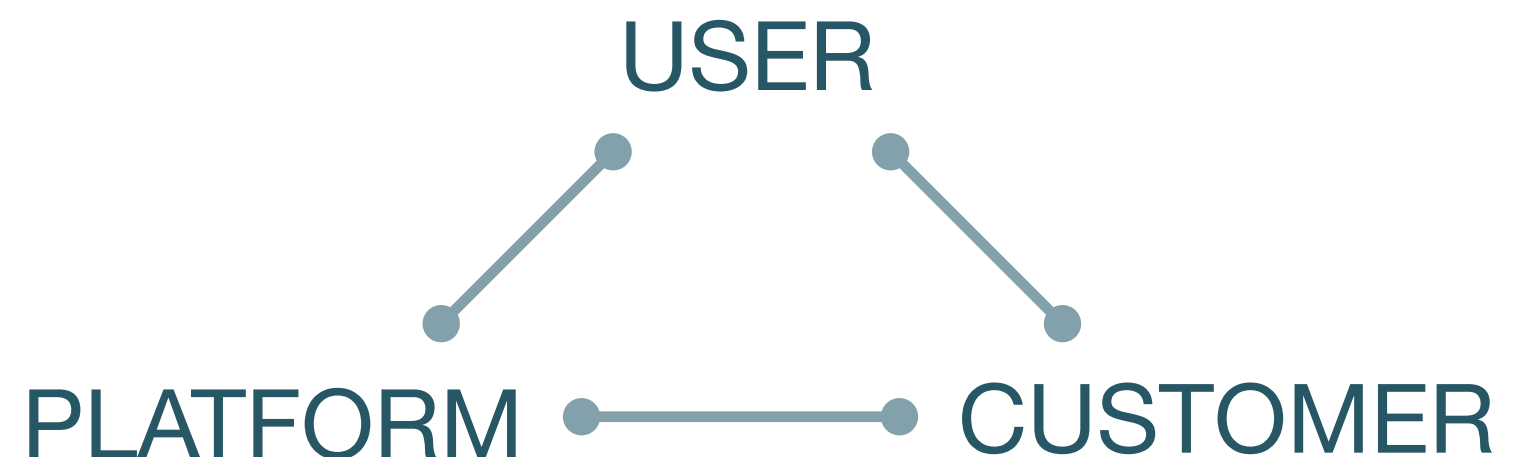
Main ethical question:

How do we balance the rights of users in relationships with platforms and companies?

What should facebook do?

What should users do?

What should US regulators do?



Stay in touch!

Data Science Community Newsletter
cds.nyu.edu/newsletter

Laura Norén
laura.noren@nyu.edu



NEW YORK UNIVERSITY