

But the IRB Approved It!

**Data Science Research Ethics
and the Challenges of Inference,
Public Data and Consent**

**Presentation for Applied Statistic Conference 2021
Slovenia (virtual)**

Jacob Metcalf, PhD

**Director, AI on the
Ground Initiative**

20.September.2021

**DATA&
SOCIETY**

Why data science is different from the perspective of research ethics

1. Norms and infrastructures for traditional research ethics are built to handle different sorts of harms, the harms of data science are rendered invisible.
2. New epistemologies: scale, speed, repurposability, predictive analytics, indefinite storage
3. Practical data ethics is developing in unexpected & informal venues.
4. Formal approaches to research ethics in data science can vary significantly across borders.

Research Ethics consists of norms & infrastructures

The shared commitments of human subjects research protections emerge out of the reaction to a series of scandals where scientists justified abuse (sometimes homicide) of individual research subjects by pointing to the general public good gained through knowledge.



The New York Times

Syphilis Victims in U.S. Study Went Untreated for 40 Years

By JEAN HELLER
The Associated Press

WASHINGTON, July 25.—For 40 years the United States Public Health Service has conducted a study in which human beings with syphilis, who were induced to serve as guinea pigs, have gone without medical treatment for the disease and a few have died of its late effects, even though an effective therapy was eventually discovered.

The study was conducted to determine from autopsies what the disease does to the human body.

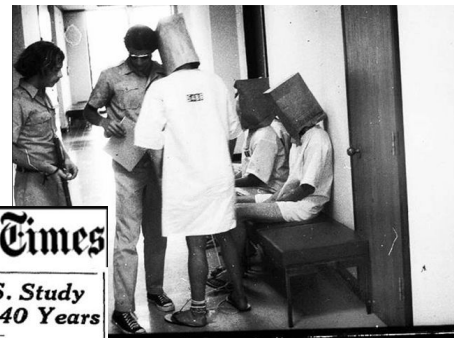
Officials of the health service who initiated the experiment have long since retired. Current officials, who say they

have serious doubts about the morality of the study, also say that it is too late to treat the syphilis in any surviving participants.

Doctors in the service say they are now rendering whatever other medical services they can give to the survivors while the study of the disease's effects continues.

Dr. Merlin K. DuVal, Assistant Secretary of Health, Education and Welfare for Health and Scientific Affairs, expressed shock on learning of the study. He said that he was making an immediate investigation.

The experiment, called the Tuskegee Study, began in 1932 with about 600 black men,



Credits:



Research Ethics consists of norms & infrastructures

1947: Nuremberg Code

1964: WMA Declaration of Helsinki

1974: US National Research Act

1976: US Belmont Report: *respect for persons, beneficence, and justice*

1981: Common Rule establishes Institutional Review Boards (IRBs) for federally funded “human subjects research”

Research: creation of new data in pursuit of generalized knowledge

Human subjects: individuals in whose lives/bodies a researcher intervenes in collection of data

2017: Major revision of Common Rule, largely addressing big data in biomedicine

European Research Ethics Committees (REC) have similar structures and ethical commitments

1947: Nuremberg Code

1953: European Convention on Human Rights

1964: WMA Declaration of Helsinki

1997: Oviedo Convention

Strong focus on biomedical research

ethics, not as much on social science

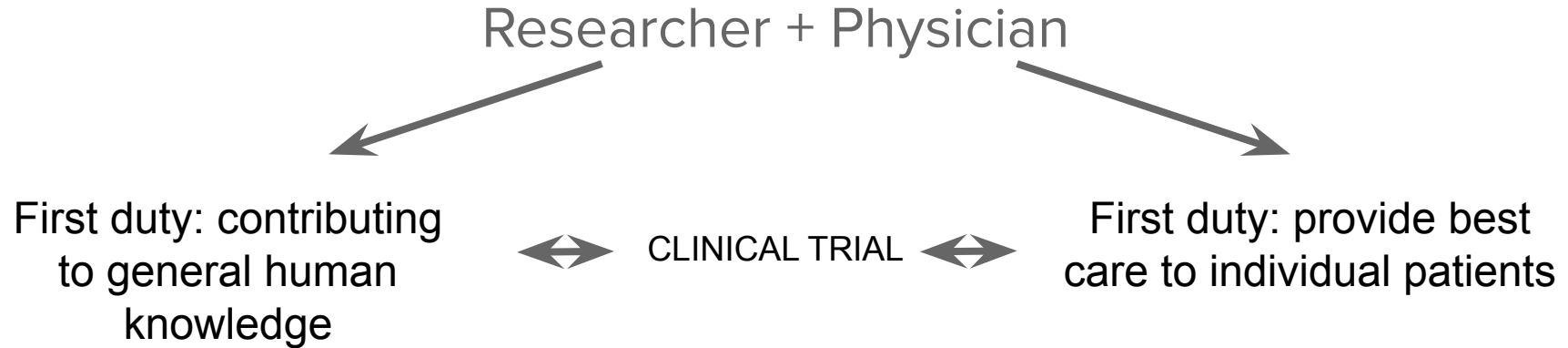
Similar definition of research, emphasizing new knowledge

Less formally attached to universities, often with nested regional structures.

Often have greater freedom to consider societal consequences.

Emphasis on consent.

Central problem research ethics tries to solve

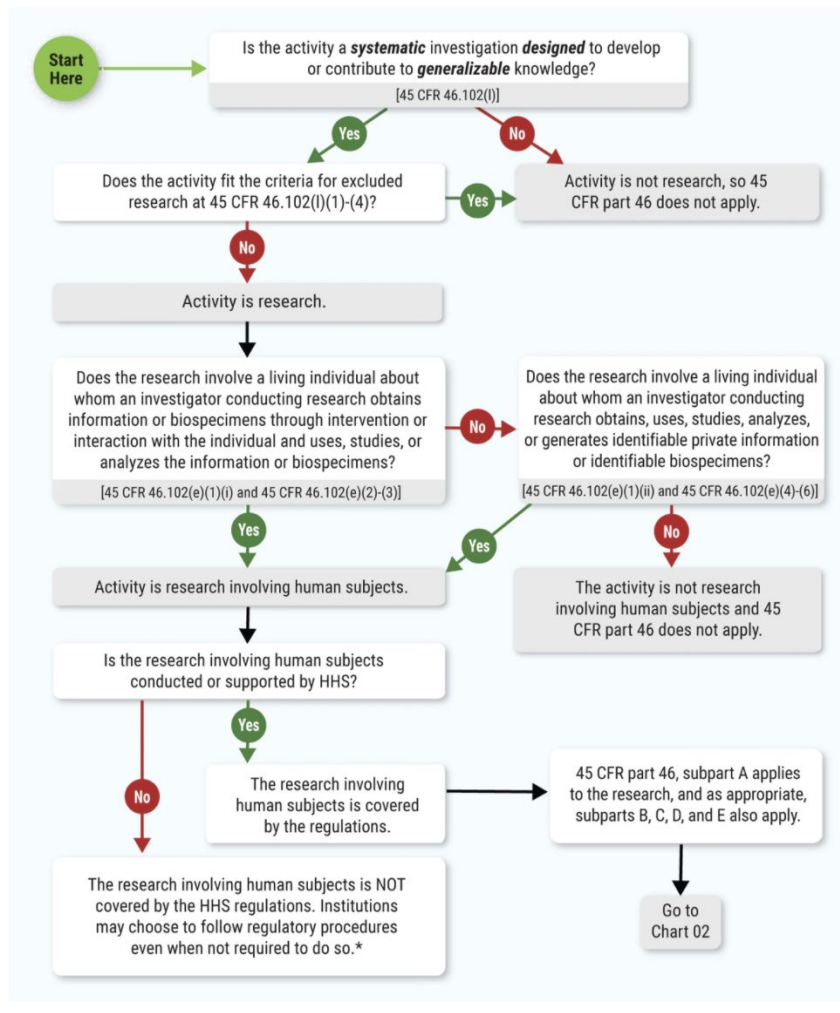


Core risk to human research subject: autonomy will be compromised and/or not receive adequate care

Common Rule is a decision tree

CR establishes Independent Review Boards as an oversight mechanism that organizations conduct themselves.

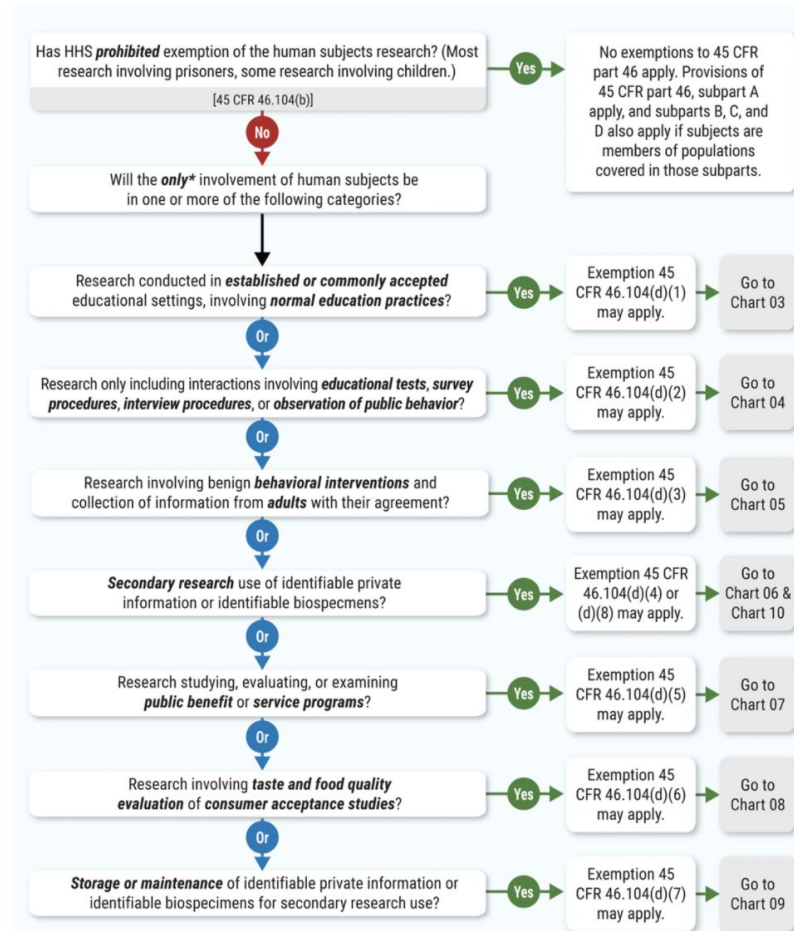
IRBs are tasked with a narrow purpose: determine if the potential harms caused to individual human research subjects by virtue of the research methods exceed every day life risks?



Common Rule is a decision tree

If a research proposal is properly determined to be a human subjects research project covered by the CR, then the IRB's next question is whether the project is "Exempt" or requires additional oversight.

Chart 02: Is the Research Involving Human Subjects Eligible for Exemption Under 45 CFR 46.104(d)?



**An “Exempt” or
“not human
subjects
research” status
is not
synonymous with
“ethical”**

- **IRB’s CANNOT consider downstream consequences**, their mandate is only for harms caused by virtue of the research methods.
- **IRB’s (mostly) CANNOT consider societal or community harms**, their mandate is for protecting individual human subjects.
- **IRB’s will not tell a research something is just a bad idea** or dubious science.
- **IRB’s will not require consultation** with communities being studied.

Data science research is almost always *technically* not human subjects research or is determined “Exempt”

- **Not research:** In many cases, data science can claim to 1) not be collecting new data, that is 2) aiming at generalizable knowledge.
- **Not *human subjects* research:** The data is just fodder for learning about statistical phenomena and technical systems, and requires no ‘intervention’.
- **Use public data:** If data are publicly accessible or observable (even by purchase, scraping or license), then they are presumed to pose no additional risk as research data.
- **Use de-identified data:** If data cannot be connected to individuals then the research is presumed to pose no additional risk.
- **Repurpose existing data:** Data science usually does not collect data, it analyzes existing data and thus has no contact/intervention with individual human subjects.

The defining feature of data science and AI/ML applications is using cheap data to infer expensive traits

- **Find cheap/accessible data about human behavior** from sources such as social media/web history/surveillance.
- **Use machine learning to find unexpected patterns** in a massive dataset.
- **Render that pattern as an algorithmic model** which can be utilized by internet platforms or in a research context.
- **Use that model to infer expensive or hard to access traits** about another group of data subjects, often in real-time.
- **This causes jumps across contexts that are ethically fraught.**

Case #1: #gayfaceAI

In 2017, Stanford researchers Michael Kosinski and Yilun Wang published a preprint paper claiming that a machine learning system could predict sexual orientation from pictures of faces better than humans could.

Artificial Intelligence

Computer Vision

AI that can determine a person's sexuality from photos shows the dark side of the data age

Posted Sep 7, 2017 by Devin Coldewey



#gayfaceAI overview

Core finding: Deep neural networks (machine learning algorithms) trained with datasets created by off-the-shelf facial recognitions software and “public datasets” perform better than humans at inferring sexual orientation from facial photographs alone.

Implied finding: that machines can do this implies there is a biological basis of sexual orientation subtly visible to trained algorithms (e.g., prenatal exposure to androgen).

Authors’ stated ethical concern: tools for intensive discrimination can be built with off-the-shelf machine learning components and widely available social data, and the public is largely unaware of this.

Public controversy: does this justify the research and the IRB’s approval?

The study relied on very narrow labels

The training set was scraped from an unnamed dating website, and mTurkers were paid to label *only* white “male” and “female” faces.

These faces were then fed to other mTurkers and the deep learning algorithm to test for the ability to estimate stated sexual preference.

Identify Adult Caucasian Males

Instructions

You will see 50 sets of 4 faces. Your job is to select **complete** faces belonging to **adult Caucasians males**. Any given set can contain between 0 to 4 adult male Caucasian faces.








You can use Back and Next button to navigate through different sets. **Please use the best of your intuition. We will carefully review the results to identify spammers.**

We welcome your feedback! There are going to be more HITs like these!

Details

1. Some images might contain a grey space on the side. It's normal and shouldn't affect your selections.
2. Some faces might be blurry. As long as you can recognize that the image represents an adult Caucasian male, the face should be accepted.
3. Faces partially covered by hats, sunglasses and hair are considered complete as long as you can recognize an adult Caucasian male.

Examples

Wrong: non-Caucasian face Black 	Wrong: non-Caucasian face clearly Latino 	Wrong: non-adult face baby 
Wrong: non-male face female-looking face 	Wrong: incomplete face part of face 	Wrong: non-human face cartoon or not a human 
Correct: Caucasian, adult, male and complete face 		

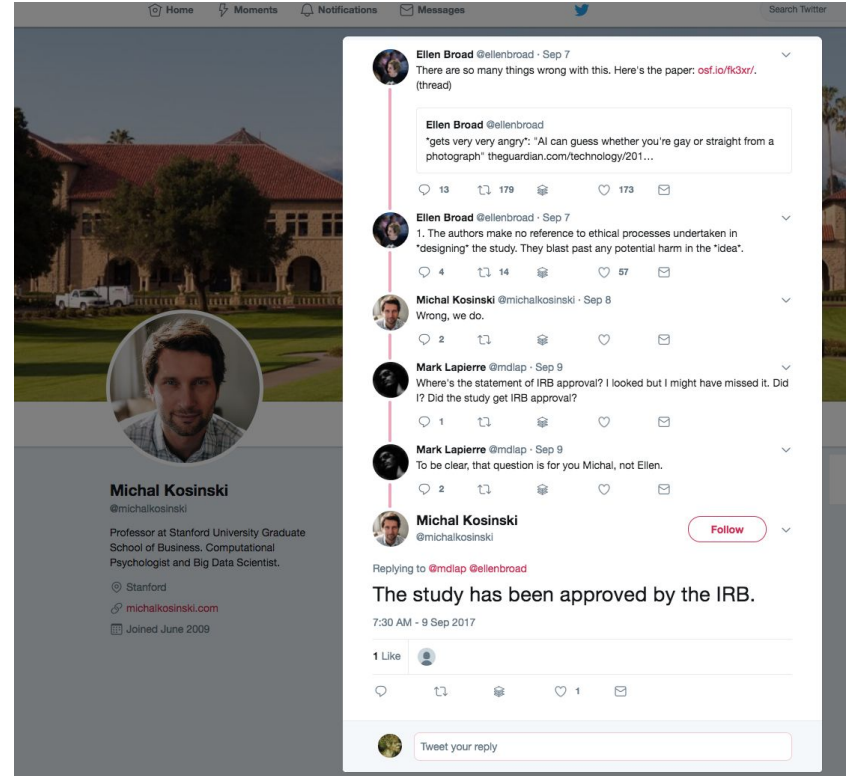
#gayfaceAI responses

A machine trained on this data could easily become a plug-and-play tool for a repressive government.

Illustrates how SBE data can be used by researchers to build tools that are turned back on a population.

But IRB's are required to *only* consider harm done to individual research subjects by virtue of research methodology.

<https://medium.com/pervade-team/the-study-has-been-approved-by-the-irb-gayface-ai-research-hype-and-the-pervasive-data-ethics-ed76171b882c>



Case #2: Cambridge Analytica as a research ethics scandal



What happens when an academic data scientist needs access to commercial “public” data?

- Without formal access to platform datasets, researchers resort to using APIs
 - But the ethical context of an API built for advertisers is significantly different than the ethical context of research
- To get “research participants”, researchers use tools that are familiar to commercial actors
 - Viral quizzes
 - MTurk
- When working with a platform, researchers often need to negotiate both a ethics review *and* a commercial data sharing arrangement

Origins of Cambridge Analytica

Timeline -- Late 2012/Early 2013: Republican post-mortem in NYC



Rebekah & Robert
Mercer



The Sea Owl,
NYC January 2013



Steve Bannon



Mark Block



Alexander Nix / SCL

Origins of Cambridge Analytica

Timeline -- Late 2012/Early 2013: viral quizzes used to infer psychological states

Private traits and attributes are predictable from digital records of human behavior



Michal Kosinski, David Stillwell, and Thore Graepel

PNAS April 9, 2013 110 (15) 5802-5805; <https://doi.org/10.1073/pnas.1218772110>

Edited by Kenneth Wachter, University of California, Berkeley, CA, and approved February 12, 2013 (received for review October 29, 2012)

Article

Figures & SI

Info & Metrics

PDF

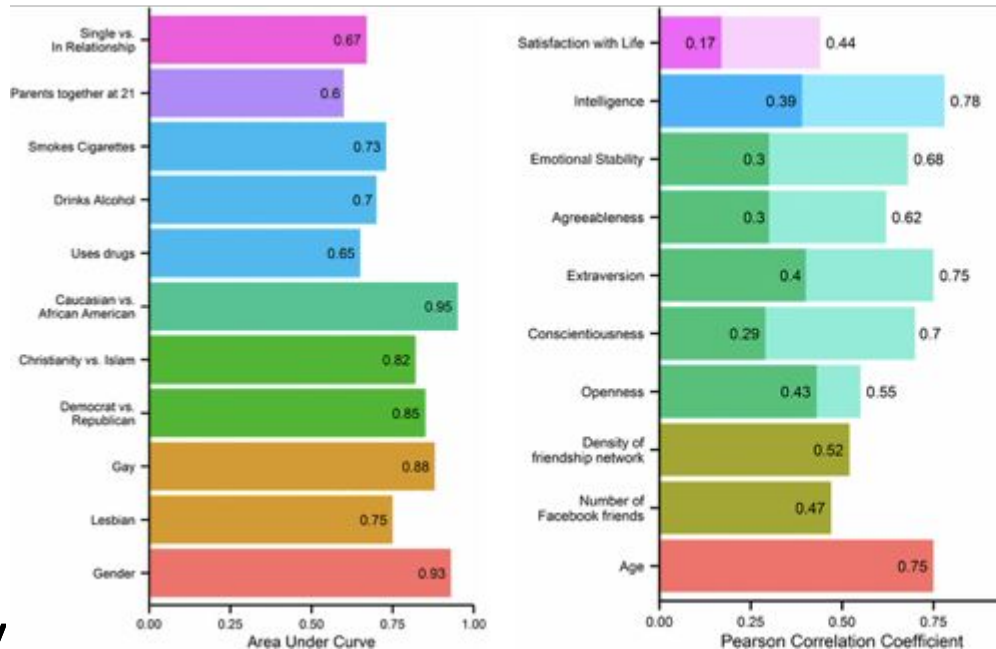
Abstract

We show that easily accessible digital records of behavior, Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender. The analysis presented is based on a dataset of over 58,000 volunteers who provided their

Origins of Cambridge Analytica

Apply Magic Sauce
PredictionAPI

Timeline -- Late 2012/Early 2013: viral quizzes used to infer psychological states



“With a mere ten ‘likes’ as input his model could appraise a person’s character better than an average coworker. With seventy, it could ‘know’ a subject better than a friend; with 150 likes, better than their parents. With 300 likes, Kosinski’s machine could predict a subject’s behavior better than their partner. With even more likes it could exceed what a person thinks they know about themselves.” Kosinski’s 2016 profile in *Das* magazine.

Origins of Cambridge Analytica

Timeline -- late 2014:
Alexsandr Kogan uses
mTurk workers to collect
data on their friends
networks by abusing FB's
commercial API and in
violation of mTurk ToS. CA
pays for this 'research' in
exchange for results.



[All Reviews](#) [Flagged By You](#) [Your Reviews](#) [Order by edit date](#)

AMT Requester	Rating [info]	Description
Aleksandr Kogan A1WU9AXHP5OUI Averages » HIT Group » Review Requester »	FAIR: NO DATA FAST: NO DATA PAY: NO DATA COMM: NO DATA	you must give an app access to your facebook account. Mar 03 2014 blue...@y... 🗨 💬 👍 👎
VIOLATES MTURK TERMS OF SERVICE [?]		
Aleksandr Kogan A1WU9AXHP5OUI Averages » HIT Group » Review Requester »	FAIR: NO DATA FAST: NO DATA PAY: NO DATA COMM: NO DATA	Violates MTurk Terms of service, this requester explicitly states "provide our app access to your Facebook so we can download some of your data—some demographic data, your likes, your friends list, whether your friends know one another, and some of your private messages. Please note that we take several precautions to ensure all of your data stays anonymous and safe, and it will only be used for research purposes. You will receive \$.50 for participating." Mar 03 2014 joysinger 🗨 💬 👍 👎
VIOLATES MTURK TERMS OF SERVICE [?]		
Aleksandr Kogan A1WU9AXHP5OUI Averages » HIT Group » Review Requester »	FAIR: NO DATA FAST: NO DATA PAY: NO DATA COMM: NO DATA	Short questionnaire - "We will then access to your Facebook to download some of your data - some demographic data, your likes, friends list, and some of your messages".. TOS violation?? Feb 25 2014 Rosey 🗨 💬 👍 👎
VIOLATES MTURK TERMS OF SERVICE [?]		
🗨 Agree, shady. I have received email requests for surveys that pay \$5.00 for such surveys, but on non-amazon site and pay with an Amazon Gift Card. They want all your Facebook info. This comment was edited by the author Fri Feb 28 20:24 PST. Feb 28 2014 jessema 🗨 💬 👍 👎		

Origins of Cambridge Analytica

Timeline -- mid-late 2014: Meanwhile, Kogan and US compatriots are seeking ethics committee approval to use this data in academic research. The Cambridge REC turns Kogan down *repeatedly*, but the UC Riverside IRB *EXEMPTS* the US collaboration *immediately*.

The researchers have made a case that Facebook allows its users to set their privacy settings to avoid being part of the study. The researchers seem to be claiming that understanding these settings is clear to all users. It is not. Even as a well-informed, long-time user of Facebook, I find the setting confusing and I know that many others do as well. Facebook's privacy policy is not sufficient protection to address my concerns.

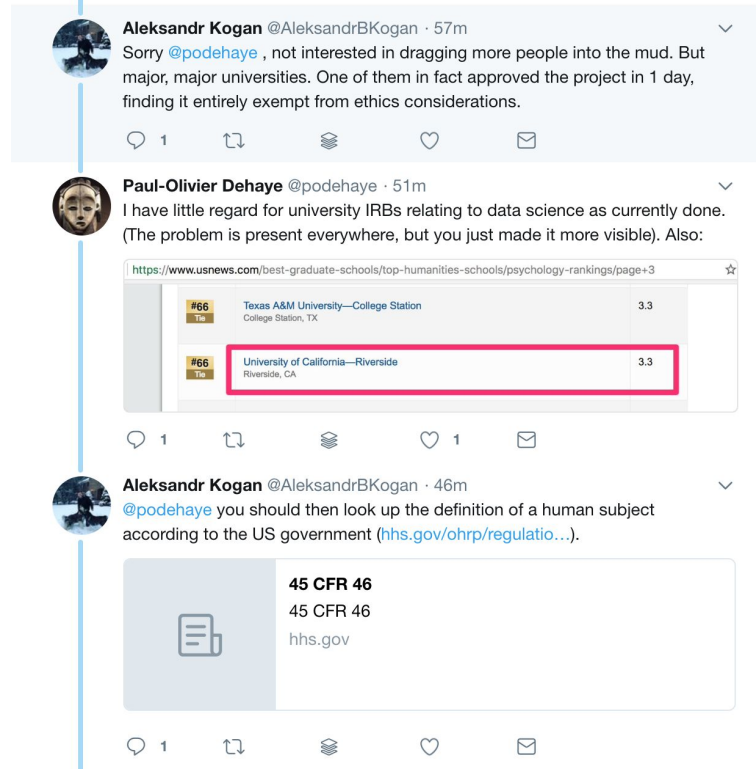
I [REDACTED] and this is the first time that I have not approved an application. I have taken a long time to return my decision and comments because I do not take this decision lightly. I have read the paper published last year in the Proceedings of the National Academy of Sciences (Kramer, Guillory, Hancock, 2014), that really kicked off discussions about using Facebook data for research. In response to that controversy, the editor of that journal expressed a concern related to Facebook's consent procedures. The editor points to Cornell University's ethical review of the project. Cornell confirmed that it did not review the project because the researchers were 'not directly

Origins of Cambridge Analytica

As far as we can tell, *no academic research* was ever published from the datasets that CA hired Kogan/GSL to collect and license. That is *perhaps* because the Cambridge REC turned down the application.

This is despite Kogan claiming to FB during a suspicious activity review that the data was being collected for academic purposes.

Using this data for commercial purposes was clearly a violation of FB's ToS.



The screenshot shows a Twitter thread with three tweets. The first tweet is from Aleksandr Kogan (@AleksandrBKogan) replying to @podehaye, stating he is not interested in dragging more people into the mud and that one major university approved the project in one day, finding it exempt from ethics considerations. The second tweet is from Paul-Olivier Dehaye (@podehaye) replying to Kogan, stating he has little regard for university IRBs and that the problem is everywhere. He includes a screenshot of a US News & World Report ranking page for psychology programs, where the University of California—Riverside is highlighted with a red box. The third tweet is from Aleksandr Kogan replying to Dehaye, stating that Dehaye should look up the definition of a human subject according to the US government, with a link to hhs.gov/ohrp/regulation... and a screenshot of the 45 CFR 46 regulation page.

Aleksandr Kogan @AleksandrBKogan · 57m
Sorry @podehaye, not interested in dragging more people into the mud. But major, major universities. One of them in fact approved the project in 1 day, finding it entirely exempt from ethics considerations.

Paul-Olivier Dehaye @podehaye · 51m
I have little regard for university IRBs relating to data science as currently done. (The problem is present everywhere, but you just made it more visible). Also:

<https://www.usnews.com/best-graduate-schools/top-humanities-schools/psychology-rankings/page+3>

Rank	School	Score
#66	Texas A&M University—College Station College Station, TX	3.3
#66	University of California—Riverside Riverside, CA	3.3

Aleksandr Kogan @AleksandrBKogan · 46m
@podehaye you should then look up the definition of a human subject according to the US government ([hhs.gov/ohrp/regulation...](https://www.hhs.gov/ohrp/regulation...)).

45 CFR 46
45 CFR 46
[hhs.gov](https://www.hhs.gov)

Case #3: Predicting Autism from family home videos on YouTube

Researchers from Keele University in the UK were interested in building AI-enabled diagnostic tools to assist/speed-up diagnosis of young children with autism.



They started with family videos of young children tagged with #autism

From there they developed a public dataset of brief clips taken from these videos which could later be used to train diagnostic tools via “digital phenotyping”. They did not ask permission or notify participants because YouTube is “public.”

Towards Automatic Screening of Typical and Atypical Behaviors in Children With Autism

Andrew Cook
School of Computing and Mathematics
Keele University
Keele, UK
a.a.cook@keele.ac.uk

Bappaditya Mandal
School of Computing and Mathematics
Keele University
Keele, UK
b.mandal@keele.ac.uk

Donna Berry
School of Psychology
Keele University
Keele, UK
d.m.berry@keele.ac.uk

Matthew Johnson
Department of Psychology
North Staffordshire Combined Healthcare NHS Trust
Stoke-On-Trent, UK
matthew.johnson2@combined.nhs.uk

II. YOUTUBE ASD DATABASE

The newly developed database presented in this work is collected from publicly available files on the YouTube video platform [1]. Initial searches were made to select a variety of short video clips highlighting stereotypical behaviors seen in children between the ages of around 3 years and 12 years with an aim of capturing a wide variety of behaviors in a variety of settings. It also presents a number of ground-truth clips of a similar number and length by which the differences in may be compared. Of particular importance to this selection process was the frame rate and resolution of the video. An attempt was made to include as wide of a range of video subjects as possible. Ground truths are created using the captions and descriptions of the user videos and expert knowledge of the authors in collaboration with an autism assessment center.

For each video a number of short sequences were chosen to highlight the key behaviours being displayed to highlight any differences between typical behaviors and the atypical actions which are the focus of this study. Each sequence of around 3 to 12 seconds was selected based upon the positioning of the subject within the frame, the presence of other subjects within

time-series data in an efficient and automated manner.

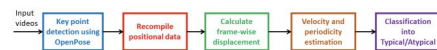


Fig. 1. Block diagram for our proposed framework for non-intrusive body parts movement analysis of typical and atypical behaviors in autistic children.



Fig. 2. Example video frames (upper row) (shown at 10 frame interval) with keypoint locations marked using OpenPose software (bottom row). Subject's face is masked for anonymity. Best viewed in color.

‘Participant’ anger contributed to withdrawal of the paper

- **Research indicates** that people do not experience social media as “public” in the same way that researchers define it (Fiesler & Proferes 2018).
- **Context matters** when engaging with social media data as a source of research data.
- **Even seemingly beneficial research can be abusive** if the people being studied aren’t consulted.

Towards Automatic Screening of Typical and Atypical Behaviors in Children With Autism

Andrew Cook, Bappaditya Mandal, Donna Berry, Matthew Johnson

This paper has been withdrawn by the authors due to insufficient or definition error(s) in the ethics approval protocol.



Lessons for IRBs & data science researchers

- **Research ethics norms are a poor fit for data science** research practices for some good historical reasons.
- **But the silence of the IRB is not a carte blanche** for or endorsement of dodgy research practices.
- **Look for alternatives venues** for research ethics support.
- **“Public” data sometimes just isn’t** actually public.
- **Research subjects want to be treated with respect** even if there 10M of them and you will never meet them.
- **Making inferences at scale about human behavior is inherently ethically risky.**

THANK YOU

jake.metcalf@datasociety.net
@undersequoias

DATA&
SOCIETY

